INVESTIGATION INTO THE MECHANISMS OF DEPRESSIVE ILLNESS

John Douglas Steele

Doctor of Medicine, University of Edinburgh

DECLARATION

I hereby declare that:

(a) this thesis has been composed by myself
(b) the work it describes was largely conducted by myself, in collaboration with a number of others from Edinburgh University: those parts of the work which were not completed entirely by myself are clearly identified in the Statement.
(c) the work has not been submitted for any other degree or professional qualification

Signed

Date

THE UNIVERSITY OF EDINBURGH

ABSTRACT OF THESIS

(Regulation 3.9.14)

Name of Candidate: Dr Douglas Steele

Address :

Postal Code:

Degree: Doctor of Medicine (MD) *Date* 6/7/2004

Title of Thesis: Investigation into the Mechanisms of Depressive Illness

No. of words in the main text of Thesis: 45,763

Functional and structural brain abnormalities have been reported in many imaging studies of depressive illness. However, the mechanisms by which these abnormalities give rise to symptoms remain unknown. The work described in this thesis focuses on such mechanisms, particularly with regard to neural predictive error signals. Recently, these signals have been reported to be present in many studies on animals and healthy humans. The central hypothesis explored in this thesis is that depressive illness comprises a disorder of associative learning. Chapter 2 reviews the brain regions frequently reported as abnormal in imaging studies of depressive illness, and the normal function of these particular brain regions. It is concluded that such regions comprise the neural substrate for associative learning and emotion. However, confidence in this conclusion is limited by considerable variability in the human imaging literature. Therefore, chapter 3 describes a meta-analysis, which tests the hypothesis that, consistent with the non-imaging literature, the ventromedial prefrontal cortex is most active during emotional experience. The results of the meta-analysis were clearly consistent with this hypothesis. Chapter 4 provides an introduction to neural predictive error signals from the general perspective of homeostatic physiological regulation. Both experimental evidence supporting the error signals, and various formal mathematical theories describing the error signals, are summarised. This provides the background to chapter 5, which describes an original fMRI study which tested the hypothesis that patients with depressive illness would exhibit abnormal predictive error signals in response to unexpected motivationally significant stimuli. Evidence of such abnormality was found. Chapter 6 describes a further original study using transcranial ultrasound and diffusion tensor imaging of the brainstem, which investigated reports of a subtle structural abnormality in depressed patients. If present, it might give rise to abnormal error signals. However, no structural abnormality was found. Finally, chapter 7 discusses the significance of these findings in the context of clinical features of depressive illness and a wide range of treatments, ranging from psychotherapy through antidepressants to physical treatments. A number of potential future studies are identified, which could clarify understanding of depressive illness.

STATEMENT

I am particularly grateful to my advisor, Prof Klaus Ebmeier, for the opportunity of working on a number of studies over the years, for funding the imaging studies described in chapters 5 and 6 from the Gordon Small Charitable Trust, and support regarding many practical issues involved in setting up and completing the imaging studies described in this thesis. Prof Eve Johnstone provided me with the opportunity of working as a Lecturer in Psychiatry at Edinburgh University. It was during this time that the studies were done. I also thank Prof Ian Reid, Aberdeen University, for allowing me to continue psychiatric imaging research on a more permanent basis, and for the time to write this thesis. The work has resulted in journal publications and book chapters. I thank Mr John Steele for carefully proof reading all of these, and this thesis.

I thank Dr Steven Lawrie, University of Edinburgh, for discussions on imaging studies reporting functional and structural abnormalities in depressive illness, for discussions on the method of selecting imaging studies used for the meta-analysis, and for independently applying the selection criteria used to identify the activation loci. Dr Martin Meyer, University of Zurich, Switzerland, helped with the IFIS programming and kept the IFIS system running near the end of the study, allowing acquisition of all the image data. Prof Joanna Wardlaw, University of Edinburgh, reported the clinical T_2 structural images and acquired the ultrasound images of the brainstem. Dr Dianna Morrison and other Royal Edinburgh Hospital consultants referred their patients for the studies. Dr Ian Marshall, University of Edinburgh, was responsible for the performance of the MRI scanner. Dr Mark Bastin, University of Edinburgh, developed the DTI pulse sequences tailored for brainstem imaging, clarified a number of issues regarding DTI, and converted the DTI images to Analyze format. Mr Tom Anderson, University of Edinburgh, helped with transfer of the ultrasound images. Dr Georg Becker, Neurologische Universitatsklinik, Wurzburg, Germany, commented on pilot ultrasound images of the midbrain obtained from two subjects during initial study planning.

My contribution to the work was as follows. I suggested that depressive illness is a disorder of associative or emotional learning. Regarding chapter 3, I had

the idea for the study, identified and obtained copies of all the included studies, applied the agreed criteria for selection of the activation loci, had the idea for the statistical analysis method, wrote the Matlab and C software for statistical analysis, analysed the data, and wrote the paper. Regarding chapter 5, I had the idea for the study, selected the game paradigm and wrote the IFIS program to implement it, recruited all the subjects and supervised their scans, had the idea for the statistical analysis method and implemented it, including the structural equation modelling technique, using Matlab and C, analyzed the data using these programs and SPM99, and wrote the paper. Regarding chapter 6, I had the idea for the study, supervised the ultrasound scanning of the same subjects as chapter 5, had the idea for the statistical analysis method and implemented it in Matlab, analysed the ultrasound and DTI data using these programs and SPM99, and wrote the paper.

The following have been published, presenting results and discussion mentioned in this thesis.

i) **Steele, J.D.** Meyer, M. Ebmeier, K.P. (2004) "Neural Predictive Error Signals and Depressive Illness", Proceedings of Human Brain Mapping Conference, Budapest

ii) **Steele J.D.**, Meyer M., Ebmeier K.P. (2004) "Neural Predictive Error Signal Correlates with Depressive Illness Severity in a Game Paradigm", NeuroImage 23, 269-280

iii) **Steele JD**, Bastin ME, Wardlaw JM and Ebmeier KP. (2004) "A transcranial ultrasound and diffusion tensor magnetic resonance imaging study of brainstem structural abnormalities in major depressive illness", Proceedings of the Tenth Annual Meeting of the British Chapter of ISMRM, P9.

iv) **Steele J.D.**, Bastin M.E., Wardlaw J.M. and Ebmeier K.P. (2005) "Possible Structural Abnormality of the Brainstem in Unipolar Depressive Illness: A Transcranial Ultrasound and Diffusion Tensor Magnetic Resonance Imaging Study" (submitted to JNNP – under review)

v) **Steele J.D**. (2005) "Depressive Illness and Emotional Learning", Current Medical Imaging Reviews (accepted for publication)

vi) **Steele, J.D.**, Glabus M.F., Shajahan P.M. and Ebmeier K.P., (1999) "Increased Cortical Inhibition in Depression: a prolonged Silent Period with Transcranial Magnetic Stimulation (TMS)" Psychological Medicine 30: 565-570

vii) **Steele J.D.** and Lawrie S.M., (2004) "Neuroimaging", Chapter 5 of "Companion to Psychiatric Studies" (2004) (ed. Johnstone E.C., Owens, D.G., et al), Churchill Livingstone, Edinburgh.

viii) **Steele J.D.** and Reid I.C., (2004) "Functional Neuroanatomy", Chapter 2 of "Companion to Psychiatric Studies" (ed. Johnstone E.C., Owens, D.G., et al), Churchill Livingstone, Edinburgh

ix) Ebmeier KP, Berge A, Semple D, Shah PJ, **Steele J.D.** (2003) "Biological treatments of mood disorders", Chapter 7 of "Mood Disorders: A Handbook of Science and Practice", (ed. Power M.), John Wiley & Sons, Ltd, New York.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

"It is important to move on from the position of defining physiological or structural [brain] abnormality in psychiatric syndromes ...the important challenge is to explore the mechanisms whereby these findings, be they physiological, structural or indeed molecular, produce the clinical features of the disorders that we seek to treat" (Johnstone, 1998). The objective of this thesis is to investigate the mechanisms of depressive illness, particularly in relation to the functional and structural abnormalities reported in imaging studies. Imaging studies are not considered in isolation, but interpreted in the context of other available evidence; e.g., studies on animals, reports of the effects of various lesions in humans, clinical features of depressive illness and various quantitative theories. As in other areas of medicine, illness is assumed to reflect disordered normal function.

The difficulties associated with such an investigation should not be underestimated. Understanding the mechanism of a disorder ultimately requires an understanding of the normal mechanism which has become disordered. Depressive illnesses comprise abnormalities of emotion, cognition and sometimes perception. Unfortunately, the neural basis of these functions is not properly understood. Indeed, it has been suggested that a consensus has begun to develop about the state of theories in neuroscience: "there aren't any, ...or at least there are not any good ones" (e.g, Eliasmith & Anderson, 2002). Neuroscience has been characterised as "data rich, but theory poor" (Churchland & Sejnowski, 1992). The situation is no better in psychology, which comprises a number of competing schools of thought, with little overall consensus (Gross, 1996).

This thesis presents three main studies, which are described in chapters 3, 5 and 6. However, they were not conceived in a theoretical vacuum. Despite the above views of neuroscience and psychology, it has nevertheless been suggested that there is reasonable agreement in one area: "the brain learns and maintains an internal model of the external world" (Barlow, 1985, p37-46; Rao, 1999). This idea is certainly not new; however, over the past decade there has been a rapid advance in

understanding the neural basis of associative learning (eg, Dayan, Kakade & Montague, 2000; Schultz & Dickinson, 2000), with parallel work in sensory (Rao & Ballard, 1999) and motor systems (eg, Baev, 1998; Wolpert, Ghahramani & Jordan, 1995). There is evidence that, during learning motivationally or emotionally significant associations, sensory perception or the performance of motor action, the brain actively predicts occurrences and compares these predictions with events, these comparisons generating predictive error signals. Predictive error signals have been extensively reported in animal studies, and over the past few years there have been many imaging reports of the same error signals in healthy humans. Numerous brain structures appear to exhibit such predictive error signals; these include the monoamine systems, which are a common site of action of otherwise diverse antidepressant drugs.

Given that there appears to be robust experimental evidence of such error signals, the question arises as to how such signals might be generated. In order to predict, it appears necessary to have a model of that which is predicted. Consequently, this has led some to suggest that the "models" of the world and self, which have been inferred to exist by psychologists, are those which give rise to error signals (e.g., Baev, Greene, Marciano, *et al*, 2002; Suri, 2001). Additionally, there appears to be an implicit assumption in much of this work that "the model" is not uniformly distributed in an amorphous fashion throughout the brain, but reflects the tendency of the brain to develop local specialisation of function, or functional segregation. Thus visual areas of the brain develop visual models (Rao & Ballard, 1999), motor areas develop motor models (Baev, 1998; Wolpert, Ghahramani & Jordan, 1995), cognitive regions develop cognitive models (Grush, 2004), and regions of the brain which are specialised for processing the motivational significance of stimuli and events develop associative learning models (Dayan, Kakade & Montague, 2000). Such models may occur in a hierarchial manner (Baev, 1998). It should be emphasised that there is no sharp cut-off between such functionally specialised regions; one region appears to merge into another.

It should be noted at the outset that although emotional and perceptual learning theories are similar to the extent that they both rely on prediction error, the frameworks are distinct. In perceptual learning theories, predictive coding is based

on a forward model of sensory inputs in which the prediction error is the difference between the input and the prediction based on the input's cause. Such prediction is usually constructed hierarchically and has been most comprehensively discussed in vision. There are similar theories of motor control. In contrast, classical temporal difference theories of emotional learning do not involve any model, and prediction refers to future outcomes in a control-theory context. Nevertheless, Suri (Suri, 2001; Suri, 2002) has recently described an internal model extension to classical temporal difference theory, allowing the formation of novel associative chains, arguing that neuronal activity may reflect the processing of an internal model during emotional learning. Critically though, prediction error in emotional learning has been linked specifically to aminergic systems and neuromodulation, whereas prediction error in perceptual theories has not.

The central hypothesis explored in this thesis is that depressive illness comprises abnormal models of the world and self reflecting disordered associative or emotional learning. The basis of this hypothesis is discussed in chapters 2 and 4. Chapter 2 begins with a discussion of the clinical features of depressive illness viewed in the light of the above hypothesis. The chapter then focuses on segregation of brain function in relation to depressive illness, beginning with a discussion of the brain regions often reported as abnormal in imaging studies. It is concluded that whilst various prefrontal and temporal brain regions are often reported as abnormal, interpretation of such findings is limited by lack of understanding of normal function. This leads to a discussion of what is known about normal function. It is concluded that in animals, these brain regions correspond to the neural substrate for associative or emotional learning, representation of the rewarding or aversive aspects of stimuli, and associated behavioural response. Imaging studies, which investigate reward based learning in humans, appear consistent. Additionally though, these brain regions *may* be most active in healthy humans experiencing emotion.

The issue of which regions of the prefrontal lobe are most active in humans experiencing emotion is addressed in chapter 3, which describes a large meta-analysis of imaging studies of emotion induction and diverse cognitive tasks, published over the past decade. The hypothesis, based on numerous non-imaging studies of animals and lesion studies in humans, was that the ventromedial prefrontal

cortex is most active during emotional states. Whilst some individual imaging studies support this hypothesis, many do not, and there is marked variability in the imaging literature. However, the findings of the meta-analysis strongly supported the initial hypothesis. Additionally, it was possible to determine centers of most likely activation for emotional and cognitive tasks, and estimates of boundaries between such regions. The former were used in a subsequent fMRI study of depressed patients.

Chapter 4 discusses various stochastic theories of brain function from the general perspective of quantitative models of physiological regulation. It begins by introducing various basic concepts central to theories of physiological regulation, such as re-entrant loop gain, goal states and error signals. This leads to a discussion of Solomon's opponent process theory of affect control, which has recently been linked to neural error signals. The prefrontal lobes comprise large re-entrant structures and a control model of their action is described and used to interpret transcranial magnetic stimulation and imaging studies of depressive illness. The remainder of the chapter focuses on the experimental evidence for neural predictive error signals in animals and humans, and summarises various formal mathematical theories which appear to describe the neural error signals. A particular focus of the discussion is the Kalman filter theory of brain function, which was used in the fMRI study of depressive illness described in chapter 5.

The study described in chapter 5 tested the hypothesis that, on the basis of numerous Kalman filter models of brain function, and extensive experimental evidence of neural predictive error signals, the brain would automatically create a prediction, approximated by a Kalman filter, of sensory input (winning or losing in a game) which would be compared with actual sensory input (actual winning or losing), to create a Kalman filter derived error signal. It was further hypothesised that patients would systematically differ from healthy controls with regard to such an error signal. The study found clear evidence of predictive error signals in healthy subjects, and these signals differed systematically in depressed patients, supporting the initial hypothesis. Additionally, an analysis of connectivity between brain regions also identified systematic differences between patient and control groups,

with regard to the error signals. This is the first study to investigate predictive error signals in a psychiatric disorder.

As stated above, the central hypothesis in this thesis is that depressive illness is a disorder of associative or emotional learning, reflected by abnormal predictive error signals. It is important to recognise that it is generally impossible to give precise definitions of psychological terms: their meaning is largely defined by the usage in the literature (Dickinson, personal communication). However, associative learning is usually taken to refer to a very general category of learning which includes emotional and reinforcement learning. It also includes learning which is purely sensory or perceptual in nature, and not directly related to reinforcement learning. Emotional learning refers to specific instances in which the target element of the association has emotional content, or is an emotional response system: often assumed to be Pavlovian in nature (Dickinson, personal communication, see also chapters 2 and 4). Consequently, with regard to the hypothesis raised in this thesis, I suggest that depressive illness is generally a disorder of emotional learning. However, in the case of *all* depressive illnesses (which include those with prominent perceptual abnormalities: i.e. psychoses) a better term may be associative learning.

Whilst the error signals have been reported in many brain regions, the monoamine systems are of particular interest due to the common mode of action of antidepressants. The nuclei of cells forming the monoamine tracts are located in the brainstem, and a series of studies have reported a subtle abnormality of the midbrain in a high percentage of unipolar depressed patients. This abnormality had been hypothesised to include the monoamine tracts, and such abnormality could potentially alter error signal function. Therefore, chapter 6 describes an attempt to replicate these findings using transcranial ultrasound and diffusion tensor imaging. Although the study was predicted to have a power of 100%, no abnormality was found; however, a possible trend in the direction of previous reports was identified. It was therefore concluded that the abnormal error signals identified in the same patients, could not be accounted for by this previously reported structural abnormality.

The studies described in chapters 5 and 6 have various limitations which are important to appreciate at the outset. The data were obtained from 15 patients with a

unipolar depressive illness of severity greater than 20 on Hamilton rating, and 14 healthy control subjects, all of whom satisfied inclusion but not exclusion criteria. The criteria are discussed in detail in chapter 5. Structured interview methods such as the Schedule for Affective Disorders and Schizophrenia (SADS) (Endicott & Spitzer, 1978), Structured Clinical Interview for DSM (SCID) or Present State Examination (PSE) (Wing, Cooper & Sartorius, 1974) were not done and it could be argued that using these methods would have improved the reliability of diagnosis over standard clinical approaches. The advantages and potential difficulties of using these scales are discussed in chapter 5. The choice of Hamilton rating scale of symptom severity also has various limitations, since it is generally regarded as biased towards biological features of depressive illness, taking less account of cognitive features and subjective distress. Alternative scales are available such as the Brief Psychiatric Rating Scale (BPRS) (Overall & Gorham, 1962) and the Montgomery-Asberg Depression Rating Scale (MADRS) (Montgomery & Asberg, 1979). The former is considered reliable but less good for mood and anxiety disorders. Only the Hamilton scale was used since it allowed direct comparison with previous studies (see chapter 6) and is by far the most commonly used scale for depression. Multiple additional scales were not used, in part because of limited resources for obtaining data, but also because multiple measures can cause difficulties in interpretation: if there is a significant imaging correlation with, e.g., the Hamilton score and BPRS, but not with the MADRS, it is difficult to interpret this finding.

Whilst two small clinical studies with the same subjects were practical and allowed concentration on patients with the most clear diagnosis, it also has a number of limitations. Sampling of the population of interest carries with it the risk of obtaining a sample that is not representative of the overall population. Indeed, just on the basis of the exclusion criteria, there are good reasons to believe that the patients who participated were not representative of the overall population of patients with depressive illness. For example, alcohol misuse is very common amongst patients with depressive illness, yet such patients were excluded since it is generally accepted that alcohol misuse can be associated with structural brain change and cause episodes of low mood.

All but one patient was receiving medication and none of the controls were. Ideally no patients would be receiving medication since there are reports of structural and functional brain change apparently as a consequence of medication, and antidepressant medication may affect the signals of interest (see chapters 5 & 7). It is possible to recruit unmedicated patients but it can take a great deal of time; e.g., in two studies in which I have recruited a total of 30 unipolar depressed patients (Steele, Glabus, Shajahan, *et al*, 2000; Steele, Meyer & Ebmeier, 2004), this taking a total of over four years (in the context of other work), only two suitable unmedicated patients were identified. Although there are ways to speed up recruitment of such patients by placing advertisements this was not done, because concern has been raised about how representative such respondents are with regard to the population of patients of interest.

Selection of suitable control subjects presents many more problems. Significant abnormality in the patient group is determined as much by the selected controls as by the patients. Controls were matched on the basis of intelligence, age and percentage of female subjects, and excluded if they had a current or possible previous psychiatric disorder, history of substance misuse, history of head injury, implanted metal, or any serious physical illness (chapter 5). Such matching allowed acquisition of controls at a rate which was practical for successful completion of the study. However it is likely that controls and patients may have differed on a number of other factors such as socio-economic status and adversity: this issue is discussed further in chapter 5.

Given such limitations, the question arises as to how the findings reported in chapters 5 and 6 should be regarded. It should not be forgotten that the initial reports of important features of an illness have often traditionally been in the form of single case reports or a small clinical series. As such, they represented only provisional hypotheses which could be tested by future studies. The work described in chapters 5 and 6 should be regarded as two small clinical studies reporting original observations requiring further investigation. Arguably, no matter how large the study, or how many times the same research group has replicated previous results (see for example chapter 6), repeated independent replication is necessary for

acceptance of findings, and this may still leave interpretation of such findings in doubt.

Finally, it should be particularly clear from the above discussion that subjects taking part in the imaging studies were not randomly selected, but were instead highly selected. This is the basis of the argument for using high statistical power fixed effects analyses (results can be interpreted only on the basis of the *sample* and not the population) *together* with the stated need for similar independent studies for generalisation to the population of interest (chapter 7). This argument is made despite the current enthusiasm for random effects analyses (which allows generalisation of sample results to the population of interest, *only* if random sampling has been undertaken, and which always has a much higher type 2 error rate). Since the study described in chapter 5 has a solid basis in animal research, and in the past few years there have been a series of independent studies on healthy human subjects reporting essentially the same signal, it is hoped that others will also investigate possible abnormality in psychiatric disorder.

CHAPTER 2

DEPRESSIVE ILLNESS AND SEGREGATION OF BRAIN FUNCTION

2.1 INTRODUCTION

This chapter begins by reviewing current concepts of depressive illness together with treatments. This is followed by a discussion of the imaging evidence for abnormal structure and function of various brain regions in depressive illness. It is concluded that whilst certain regions, which include the anterior cingulate, orbitofrontal cortex and basal ganglia, are repeatedly reported as being abnormal, interpretation of such findings is substantially limited by lack of understanding of normal function. The following section therefore summarises existing knowledge of the normal function of these brain regions, concluding that in animals they correspond to the neural substrate for associative or emotional learning, representation of the rewarding or aversive aspects of stimuli, and associated behavioural response. Imaging studies, which investigate reward based learning in humans, appear consistent. Additionally though, these brain regions are active in healthy humans experiencing emotion. The link between normal human emotions and reinforcers is then discussed, followed by a statement of various hypotheses.

2.2 CLINICAL FEATURES OF DEPRESSIVE ILLNESS

" 'Major depression' is a misnomer in the sense that anxiety symptoms, irritability and hypohedonia (diminished ability to find interest or reward in previously motivating activities) are endorsed at least as commonly by subjects meeting criteria for MDD [major depressive disorder] than 'depressed mood' itself (Drevets & Todd, 1997). Food, sex, hobbies, social behaviour and work accomplishments are no longer rewarding, and depressives *(sic)* become inactive because of the inability to motivate themselves to engage in such behaviours. Social activity is avoided because social contacts become anxiety provoking and minor stressors overwhelming. Panic attacks occur in one-third of cases of MDD and up to one-half of cases of BD [bipolar disorder]. These emotional symptoms are

accompanied by prominent fatigue, psychomotor slowing or agitation, and the perception of 'psychic pain' (Carroll, 1994). The psychological manifestations include preoccupation with death, suicide, guilt, self deprecation and hopelessness. About 15% of patients hospitalised for a MDE [major depressive episode] eventually die by suicide, and about one half of all suicides occur in the context of a MDE (Drevets & Todd, 1997)" (Drevets, 1999).

Anhedonia is a core feature, and it would be difficult to imagine diagnosing the above syndrome as depression if the single feature of anhedonia were absent, and the patient continued to enjoy activities (as much as would be practical, given the aversive experiences). Other common features, such as loss of appetite and consequent weight loss, are also directly linked to hedonic mechanisms, and may be consistent with the hypothesis of depression being a disorder of associative learning. The aversive and distressing aspects of depressive illness are predominately caused by the anxiety component, often underestimated unless expressed as agitation, and persistent cognitive features, such as guilt, self-deprecation and hopelessness, together with their emotional accompaniments. The frequent clinical association of low mood with prominent anxiety is consistent with a shift in emotional responsiveness, suggesting that underactivity of the reward system often occurs with overactivity of the aversive system (see section 2.4.1). The latter might also be reflected by an endocrine stress response (such as hyper-cortisolism) and the characteristic sleep disturbances, which often occur in the illness. Of course, severely unwell patients may additionally become psychotic and experience mood congruent hallucinations, delusions, ideas of reference and passivity phenomena. Substance misuse, most commonly of alcohol, frequently complicates the natural course and treatment of the illness, and it is interesting that the common site of action of misused drugs, including alcohol (Bardo, 1998), corresponds to regions reported as abnormal in *non*-substance misusing depressed individuals, suggesting a further link between hypothesised disordered brain reward and aversion mechanisms and mood disorder.

Any comprehensive and testable theory of the mechanisms of depressive illness should address associated existing theories of illness. The most influential theory, with the widest empirical support, relates to the common mode of action of

antidepressants. The monoamine hypothesis of mood disorder is based on the observation that virtually all empirically discovered effective antidepressants, which are otherwise quite diverse chemicals, have the common action achieved by different means, of increasing the levels of monoamines such as serotonin and noradrenaline in the brain (Bloom & Kupfer, 1995; Schildkraut, 1965). There is some evidence from animal studies of dopamine enhancement as an effect of antidepressants (e.g., Bonhomme & Esposito, 1998) and some clinical evidence of dopamine hypofunction (Tremblay, Naranjo, Cardenas, *et al*, 2002). Ebmeier and Ebert have reviewed early imaging evidence for a dopaminergic disturbance in depressive illness (Ebmeier & Ebert, 1996). Nevertheless, persistent use of non-antidepressant drugs which specifically release dopamine (e.g., stimulants, opiates, and to some extent alcohol) are associated with reports of an increased prevalence of depressive illness (McIntosh & Ritson, 2001). A possible resolution to this apparent paradox is discussed in chapter 5. Crucially though, the fact that the normal action of monoamines such as serotonin and dopamine is to modulate associative or emotional learning, and that all antidepressants have direct effects on these systems, is further evidence that depressive illness may consist of a disorder of associative learning.

The other main group of existing theories of depressive illness are psychological, and are also based on treatments for depressive illness. The most influential is cognitive-behavioural therapy which focuses on "negative automatic thoughts", seeks to challenge these thoughts by looking for alternative explanations, then advocates practising such "alternative thinking" (e.g., Hawton, Salkoviskis, Kirk, *et al*, 1994). Interpersonal psychotherapy is unlike cognitive behavioural therapy, and views dysfunctional interpersonal relationships as a cause or maintaining factor in depressive illness, seeking to treat the illness by resolving the interpersonal problems (Klerman & Weissman, 1984). Another traditional approach is psychodynamic. Whilst there is some evidence for psychological approaches, the evidence for efficacy and safety is less than that for drug treatment (Ebmeier, Berge, Semple, *et al*, 2004), which may accord with practical clinical experience.

Clearly there is a huge gap between the monoamine theory (which does not explain *how* increased serotonergic and noradrenergic levels effectively treats depressive illness, nor why there is a delay in response, and why misuse of dopamine

releasing substances may make things worse in the long run) and psychological theories (which recognise a mind but not a neural substrate, and cannot account for the effectiveness of antidepressants). Depressed patients clearly express their illness in psychological (e.g., negative automatic thoughts) and social (e.g., dysfunctional interpersonal relationships) terms. Consequently, regardless of the relative practical efficacy of psychological and drug treatments, there is clearly a need to develop a testable theory of the mechanisms of depressive illness, which can begin to simultaneously address the empirical evidence for psychological approaches, the mechanism of action of antidepressants, the objective clinical features of depressive illness, and the subjective experience of the illness. Of clear interest then is the mechanism by which a systematic and stereotypical *bias in the interpretation of information* (O'Carroll, 2004, p138) occurs, manifested clinically by "negative automatic thoughts" such as inappropriate guilt, self deprecation and hopelessness.

Although arguments can be made against the *validity* of the concept of depressive disorder, there is little doubt about its clinical *utility* (Kendell, Lawrie & Johnstone, 2004). The current ICD10 or DSM IV classification allows identification and communication about patients with similar groups of symptoms and signs, common responses to a class of medications (which have a common action of increasing monoamine levels), and a similar long term outlook. However, it is possible that the *mechanisms* by which psychotic symptoms occur in unipolar depression, bipolar mood disorder, substance misuse and schizophrenia are identical, and similarly that the mechanisms by which depressive symptoms occur in unipolar depression, bipolar mood disorder, substance misuse and schizophrenia are also identical. The efficacy of two different classes of medication on these syndromes, regardless of diagnosis, would appear to support this suggestion. This suggests that focusing on one diagnostic category may be too restrictive; however, alternative approaches are not common, and are likely to result in substantial problems of interpretation. Discussion of the merits and difficulties associated with a categorical versus dimensional approach to clinical and research work can be found elsewhere (Kendell, Lawrie & Johnstone, 2004).

Another important issue concerns "trait versus state". It has become common to distinguish between these concepts: e.g., state anxiety can be defined as an

unpleasant emotional arousal occurring in the context of threatening demands or dangers. Trait anxiety is defined by contrast as stable individual differences in the tendency to respond with state anxiety to the context of threatening circumstances. Similar definitions can be constructed for state and trait depression, with "loss" substituted for "threat", though such definitions clearly overlap (i.e., a loss can also be regarded as a threat and *vice-versa*): this is reflected by very similar treatments for both disorders. In genetics research, the term trait refers to any genetically determined characteristic amenable to a segregation analysis. As noted by Goodwin, the identification of "neuroticism" as a personality trait (as defined by Eysenck) may be relevant to the understanding of depressive illness (Goodwin, 1998), since Gray specifically proposed that neuroticism reflects an enhanced trait *sensitivity* of the "behavioural inhibition system" (see section 2.4.1) to negative experiences (Gray & McNaughton, 2000). There is experimental evidence that high trait neuroticism individuals show a greater sensitivity to negative mood induction, and independently that high trait extraversion individuals show a greater sensitivity to positive mood induction (Goodwin, 1998). Gray elaborated this into a graphical representation of a two-factor version of reinforcement (reward vs. punishment) sensitivity theory, in which his personality axes (anxiety and impulsivity) are shown as being at 45 degrees to Eysenck's (neuroticism and introversion) (Gray & McNaughton, 2000). It is possible that high trait neuroticism individuals with enhanced sensitivity of the behavioural inhibition system, in the context of sustained environmental or physiological "stress", would be predisposed to develop the failure of emotional homeostasis that Mayberg and others have suggested constitutes a depressive illness (see section 5.4.1). Long term, but not short term, antidepressant use may have desensitising effects (see section 5.4.2.1).

2.3 ABNORMAL BRAIN REGIONS IN DEPRESSIVE ILLNESS

Imaging studies provide by far the most substantial evidence for abnormalities of brain structure and function in depressive illness. Although a neglected topic of research (Jeste, Lohr & Goodwin, 1988), there is significant evidence of structural abnormalities in the brains of patients with unipolar and bipolar mood disorders. Only a brief discussion of the main findings can be

presented here. For more detailed discussion on some of these, and many other issues, see Ebmeier's review (Ebmeier & Kronhaus, 2002).

2.3.1 General Structural Abnormalities

A large meta-analysis concluded that radiological signs of ventricular enlargement and sulcal prominence in unipolar and bipolar mood disorder patients were highly significant with an effect size only slightly less than in patients with schizophrenia (Elkis, Friedman, Wise, *et al*, 1995). Similarly, another review concluded that there was clear evidence of cerebral and also cerebellar atrophy (Videbech, 1997). Signal hyperintensities are punctuate lesions with reduced myelination and neuropil atrophy, which are best visualised on T_2 MRI images. When coalescent, these lesions are also known as subcortical leucoencephalopathy. A recent meta-analysis concluded that such hyperintensities occurred more frequently than expected in both unipolar and bipolar disorder (Videbech, 1997). Late onset unipolar patients, in particular, tend to have many lesions, mostly periventricularly but also in the thalamus, in deep white matter and basal ganglia. This pattern suggests dysfunction in the prefrontal basal ganglia thalamocortical loops (see below).

2.3.2 Localised Structural Abnormalities

Prefrontal Cortex Studies have reported generalised reduction in prefrontal lobe volume in patients with major depression (Videbech, 1997). Specific reductions in the grey matter of the subgenual anterior cingulate have been reported (Drevets, 2000a), and in the medial orbitofrontal gyrus of patients with mood disorder. Post-mortem studies have also reported reduction in grey matter glia and neurones in the subgenual anterior cingulate and orbitofrontal cortex (Drevets, 2000a). Imaging studies reporting rostral anterior cingulate grey matter loss are not specific to depression. Several volumetric studies of schizophrenia have found such differences (Job, Whalley, McConnell, *et al*, 2002). It is noteworthy that Drevets and colleagues specifically recruited subjects with strong family histories of mood disorders, and all these studies excluded confounders such as alcohol misuse, which can be associated with structural brain abnormality.

Temporal Lobe A few studies have reported generalised reduction in temporal lobe volume, although others have not found this (Videbech, 1997). There have been some reports of a specific reduction in hippocampal volume in mood disorder patients (Shah, Ebmeier, Glabus, *et al*, 1998; Sheline, Wang, Gado, *et al*, 1996). In the first study, it was found that such changes occurred only in the most treatment resistant patients. In the second study, volume reduction correlated with the duration of illness. Following animal work, it has been speculated that these changes may relate to hypercortisolaemia. Some studies have also reported increased volume of the amygdala in depressive illness. Volume reductions in the medial temporal lobe are not specific to mood disorder, and are also reported in studies of schizophrenia.

Basal Forebrain Several studies involving patients with unipolar depression have reported reduction in basal ganglia volume. This is in contrast to studies of bipolar disorder, where either no change in size has been reported, or increased size has been found (Videbech, 1997). In studies of patients with schizophrenia, increased basal ganglia volume, apparently secondary to neuroleptic treatment, is typical. Such an effect may also confound studies of bipolar mood disorder.

Brainstem Significant reductions in brainstem, cerebellar vermis and medulla size occurring in patients with depressive illness are occasionally reported. A common mode of action of virtually all empirically derived antidepressants is to increase serotonin and noradrenaline levels, and to a lesser extent dopamine. Antidopaminergic neuroleptics are often used to treat mania. The nuclei supplying these monoamines to the rest of the brain are located in the midbrain. Of interest then is a series of studies using transcranial ultrasound and MRI in unipolar depressed patients, reporting a structural abnormality in the midbrain raphe (Becker, Berg, Lesch, *et al*, 2001). It has been argued that the imaging abnormalities in unipolar mood disorder are consistent with a relative loss of medial forebrain bundle fibres: chapter 6 describes an attempt to replicate these findings.

2.3.3 General Functional Abnormalities

Many functional imaging studies of depression have been undertaken over the past decade. Imaging techniques have improved greatly over the same period.

One noticeable aspect of early work is that the spatial resolution of such images was relatively poor, and often there was only a limited attempt to localise abnormal findings, such that reports of "prefrontal hypoactivity" (or hyperactivity) were not uncommon. Despite this, reviews of a number of imaging studies, which did achieve better localisation of functional abnormality, are available (Drevets, 2000a). Studies of the resting state (i.e. without a task being done by the subjects during scanning) tend to report dorsolateral hypoactivity, and ventromedial hyperactivity. More recent studies often focus on abnormalities within specific brain regions: these will now be discussed.

2.3.4 Localised Functional Abnormalities

Ventromedial prefrontal area This is taken to mean the rostral anterior cingulate and orbitofrontal cortex, although the boundaries of the area are unclear. The reason the boundaries are unclear is because there is no consensus in the literature. Undoubtedly it includes the medial orbitofrontal cortex and the adjacent subgenual anterior cingulate. It may include part of the lateral orbitofrontal cortex and some of the rostral anterior cingulate. Part of the motivation for the sterotactic meta-analysis described in chapter 3 was to try to clarify where the boundaries should best be drawn: see figure 3.5 and associated discussion.

A number of studies have reported abnormal metabolic activity in the subgenual anterior cingulate and adjacent supragenual anterior cingulate in depressive illness (Drevets, 2000a). Such increased regional activity is not specific to depressive illness, occurring with numerous emotion induction imaging paradigms (Bush, Whalen, Rosen, *et al*, 1998), and in various anxiety disorders. However, two studies have reported right subgenual anterior cingulate activity correlating with depression severity (Drevets, 2000a). Abnormal activity of the orbitofrontal cortex is also frequently reported, but is not specific to mood disorder, occurs in various anxiety disorders, and normally on tasks involving a combination of emotional and cognitive processing e.g. gambling tasks.

Dorsolateral prefrontal area This is taken to mean the dorsolateral prefrontal cortex and caudal anterior cingulate, although the boundaries of the area are again unclear. The region, whilst often appearing hypoactive in resting studies of

depressive illness, tends to be active when subjects engage in diverse cognitive-attentional tasks (Bush, Whalen, Rosen, *et al*, 1998). Dorsolateral hypoactivity is again not specific to depressive disorder, also being reported in schizophrenia and being associated with negative symptoms. It has been suggested that reciprocal activation-deactivation in brain regions involved in emotional and cognitive function is a general finding in imaging studies, whether subjects are depressed or not (Drevets & Raichle, 1998). This suggests the existence of a normal mechanism by which strong emotion, such as severe depression or anxiety, might interfere with cognitive function in a reversible manner. It also suggests a means by which cognitive methods (i.e. distraction) might reduce experience of unpleasant emotion (Drevets & Raichle, 1998).

Temporal lobe Higher resting amygdala activity in unipolar and bipolar mood disorder patients, relative to controls, has been reported which correlates with depression severity (Drevets, 2000a). Antidepressant treatment has been found to reduce amygdala activity in animals and humans. Medicated mood disorder patients in remission, who relapse on tryptophan depletion, have been reported to have higher resting amygdala activity. High resting amygdala activity may be specific to mood disorder and linked to hypercortisolism (Drevets, personal communication, Drevets, Gautier, Price, *et al*, 2001). Increased cortisol may influence the limbic BGTCL (section 2.4.2.4) through its action on amygdala activation and ventromedial prefrontal regions (Erickson, Drevets & Schulkin, 2003).

Considerable experimental evidence in animals and humans links the amygdala with processing emotion related information including facial expression (Aggleton, 2000). It has been reported that unmedicated euthymic women with a history of major depression have an enhanced recognition of fearful facial expression compared to controls, but this normalises following acute citalopram infusion (Bhagwagar, Cowen, Goodwin, *et al*, 2004). Additionally, increased positive vs. negative emotional perception and memory in healthy volunteers has been reported following selective serotonin and noradrenalin reuptake inhibition (Harmer, Shelley, Cowen, *et al*, 2004). This has lead Goodwin and colleagues to suggest a theory of antidepressant drug action focusing on alteration of emotional bias (Harmer, Hill, Taylor, *et al*, 2003). Previously it has been reported that amygdala activation in

response to sad facial expressions differs in depressive illness: habituation takes longer than in matched controls (Drevets, Price, Bardgett, *et al*, 2002). Further imaging investigation into the effects of acute versus long term antidepressant treatment, in relation to amygdala function and using ambiguous emotional target stimuli, is clearly of interest and therefore planned. The normal function of the amygdala is discussed in section 2.4.3.2.

Basal forebrain Reports of abnormal activity in basal ganglia structures also occur but appear somewhat inconsistent. Abnormal thalamic activity has been reported on several occasions. It has been speculated (Drevets, 2000a) that such cortical and subcortical abnormalities represent dysfunction in basal ganglia thalamocortical re-entrant loops (Alexander, Crutcher & DeLong, 1990). The motor loop is believed to be functionally abnormal in Parkinson's and Huntington's diseases (Alexander, Crutcher & DeLong, 1990) and the limbic loop may be abnormal in an analogous manner in mood disorder.

2.3.5 Summary

A number of brain regions have been repeatedly reported as exhibiting abnormal structure and function in depressive illness; i.e., anterior cingulate, orbitofrontal cortex, dorsolateral cortex, amgydala and hippocampus, plus subcortical structures such as the basal ganglia and thalamus. Unfortunately, interpretation of such studies is substantially limited by lack of understanding of the normal function of these brain regions. Consequently, this topic is discussed in the next section.

2.4 NORMAL FUNCTION OF ABNORMAL BRAIN REGIONS

2.4.1 Representation of Rewarding and Aversive Aspects of Stimuli

In this section, many references are made to brain regions "representing" objects; e.g., events, reward, emotion or cognition. Understanding representation is a central goal in neuroscience (Eliasmith & Anderson, 2002, p4). Direct observation of neuronal firing in single unit recording, or collectively and indirectly via imaging, contingent on external events, suggests representation. If collective firing is localised to a brain region, this may be evidence of segregation (specialisation) of

function. As will be discussed, the existing evidence, obtained from extensive studies on animals and recent imaging studies on healthy humans, clearly suggests that specific brain regions, and not all brain regions, are involved in the representation of the rewarding and aversive qualities of stimuli.

Brain *reward* mechanisms have been investigated since the 1950s using direct electrical stimulation via implanted electrodes, and the injection of drugs using microcannula techniques (e.g., Rolls, 1999). Figure 2.1 summarises the results of many studies on animals investigating the site of action, indicated by successful reward conditioning, by chemically diverse drugs. The ventral tegmental region is part of the brainstem (figure 2.8), and the ventral striatum is easily located (figure 2.1). Direct electrical stimulation of these regions can produce similar behavioural conditioning effects (Rolls, 1999). All the drugs mentioned in figure 2.1 are misused by humans because of their common action of producing a transient pleasurable mood state. Imaging studies of humans addicted to a number of these substances have reported brain activity which appears consistent with animal studies (e.g., Steele & Lawrie, 2004a). It has been suggested that these substances are rewarding *because* they artificially activate brain systems which have evolved to control appetitive behaviour (e.g., feeding, drinking, sexual activity) (see for extensive discussion and references, Rolls, 1999). Collectively, some of these sites (medial prefrontal cortex, nucleus accumbens, ventral pallidum) are components of the limbic basal ganglia thalamocortical re-entrant loop (see below). Clinically, heavy repeated use of all these substances is associated with an increased prevalence of depressive illness (McIntosh & Ritson, 2001), and substance misuse often complicates the treatment of depressive illness in general. Since the brain regions affected by these substances have often been reported to exhibit abnormal function and sometimes structure in studies of depressive illness, present in highly selected subjects who are *not* misusing such substances, this might also be consistent with the hypothesis that depressive illness comprises a disorder of associative learning.

In contrast to studies on reward, decades of animal studies on *aversive* learning (i.e., fear conditioning) have implicated overlapping but, to some extent, different brain structures (Gray & McNaughton, 2000, figure 1.8). Gray's "behavioural defence system" is shown in figure 2.2.

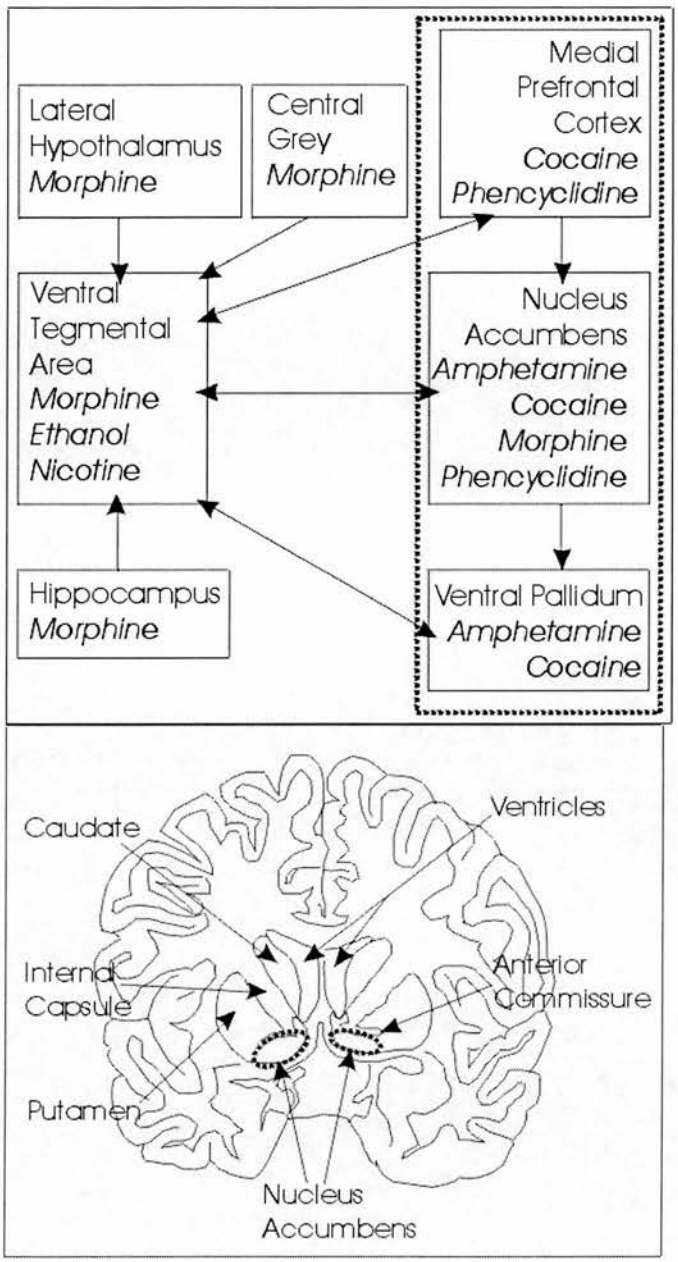


Figure 2.1 The top diagram shows a summary of brain regions and drugs supporting associative learning in animals (adapted from, McBride, Murphy & Ikemoto, 1999). The regions enclosed in the dotted rectangle are components of the limbic basal ganglia thalamocortical loop (see below). The bottom diagram shows the location of the nucleus accumbens which is found at the infero-medial boundary of the internal capsule: the ventral pallidum is much smaller (e.g., Mai, Assheuer & Paxinos, 1998) and so is difficult to identify in human imaging studies (see also chapter 5).

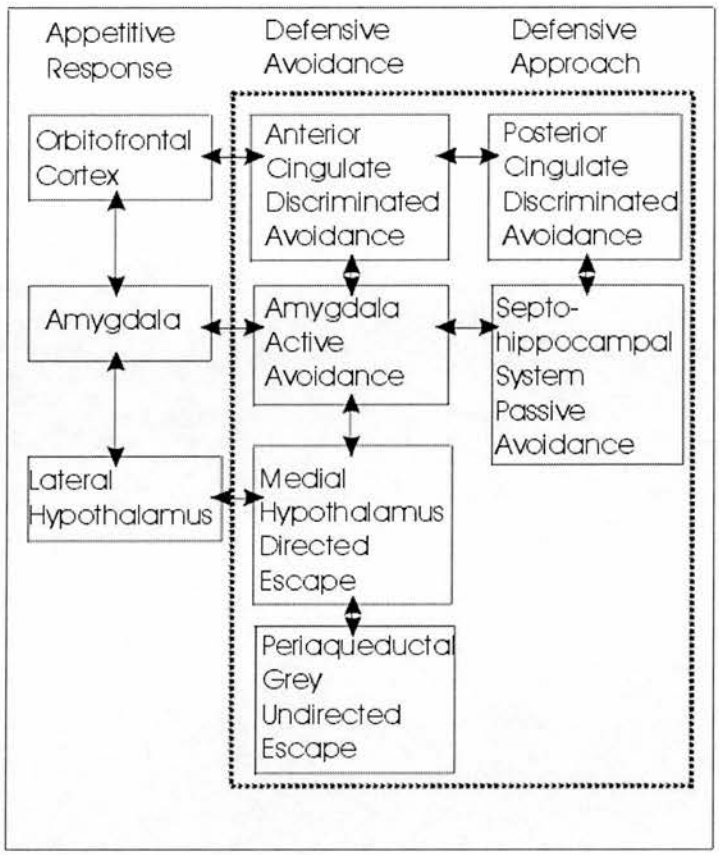


Figure 2.2 Gray and McNaughton's "behavioural defence system" (Gray, 1995) is within the dotted rectangle. Brain regions supporting more complex behaviours are shown at the top of the hierarchy. A suggested hierarchy of brain regions supporting appetitive behaviour is shown to the left of Gray's model. The amygdala is involved in appetitive responses as well as aversive responses (Cardinal, Parkinson, Hall, *et al*, 2002; LeDoux, 1998a), but the flexibility of behavioural response appears less than that supported by the orbitofrontal cortex (LeDoux, 1998a; LeDoux, 1998b; Rolls, 1999). The medial forebrain bundle runs through the lateral hypothalamus: both structures are clearly associated with simple appetitive behavioural responses (Rolls, 1999). The diagram is a considerable simplification; e.g., since the orbitofrontal cortex also supports aversive representation, and the anterior cingulate can be activated with pleasant stimuli.

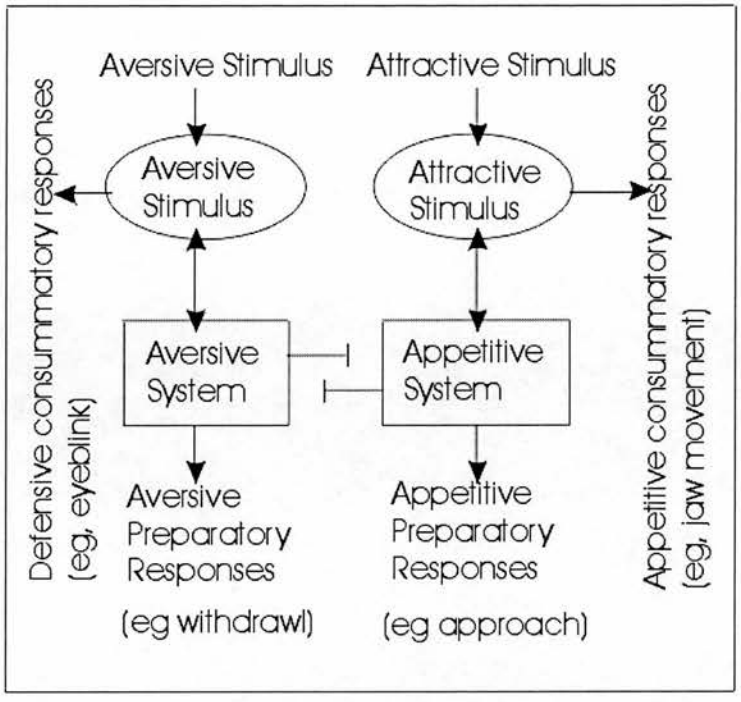


Figure 2.3 Dickinson and colleagues model of two brain motivational systems (Dickinson & Balleine, 2002; Dickinson & Dearing, 1979) which is similar to an earlier formulation (Konorski, 1967). This representation is simplified from the original, in that the appetitive and aversive stimuli include both unconditioned and conditioned forms of the stimuli. Mutual inhibitory interactions exist between the aversive and appetitive systems (c.f., the two divisions of the peripheral autonomic system).

This theory is largely consistent with other formulations (Davis, Rainnie & Cassell, 1994; Graeff, 1993; LeDoux, 1998a). Although there is much speculation in the imaging literature about brain regions specifically associated with reward and aversion, or elevated mood states versus depressed mood states (or a "depression region", such as the subgenual anterior cingulate), the evidence for such separation is inconclusive. Similarly, the much older literature on animal studies, which clearly identified specific brain regions supporting behavioural conditioning, did not identify any of the regions as invariably and exclusively associated with either reward or aversion.

Considering interactions between the systems, it is necessary to mention the experimental evidence for opponent interactions between serotonin and dopamine (Daw, Kakade & Dayan, 2002), since this forms a link between associative learning and depressive illness. There is unequivocal evidence of dopamine being involved in appetitive conditioning (e.g., Rolls, 1999). Despite the large number of effects of serotonin, in animal and human studies it has been particularly associated with aversion, punishment and anxiety (Graeff, 1993; Harrison-Read, Tyrer & Sharpe, 2004, p460). The *behavioural* actions of serotonin are in opposition to dopamine: serotonin appears to act by antagonising dopamine function at the level of the ventral tegmental area, substantia nigra, nucleus accumbens and striatum (see for references, Daw, Kakade & Dayan, 2002). Additionally, this opponency appears asymmetric: there is less evidence of dopaminergic inhibition of serotonergic action than *vice-versa* (Daw, Kakade & Dayan, 2002). Other reviews consider the evidence for the available data supporting one appetitive and one aversive system (perhaps as figure 2.2), rather than multiple appetitive and aversive systems, or one single combined system (Dickinson & Dearing, 1979): figure 2.3. Simplistic though it may be, it is tempting to speculate that depressive illness involves underactivity of the appetitive system and overactivity of the aversive system.

Although not directly related to current concepts of neural representation, it is necessary to also mention different neurotransmitter system receptors. It should not be thought that the above concepts of the reward and aversion systems ignore the large number of apparently distinct receptor subtypes identified by molecular biologists. A useful analogy can be drawn with the opposed actions of the peripheral

sympathetic and parasympathetic nervous system. Although this concept was originally proposed by Cannon in the 1930s, it remains relatively unchallenged, and the similarly large number of receptor subtypes which have subsequently been identified are usefully understood in the context of Cannon's theory.

Barto has reviewed an array of different neurotransmitters and receptor subtypes in relation to reinforcement conditioning studies (Bardo, 1998). There are a significant number of different receptor subtypes: the serotonergic system alone is currently believed to have 14 (Higgins & Fletcher, 2003). Consequently, it will only be possible to focus on a few receptors of particular interest. Higgins and colleagues noted that despite some inconsistencies in the effects of serotonergic manipulations, "there is a consensus between different laboratories using different self-administered drugs, and different schedules of reinforcement, that generalized elevations in 5-HT neurotransmission appear to reduce cocaine and amphetamine self-administration" (Higgins & Fletcher, 2003, p154). They go on to argue that the 5-HT_{2C} receptor is particularly important in this regard. Deakin and Graeff attempted to reconcile the "5-HT excess" hypothesis of anxiety with the "5-HT deficiency" hypothesis of depression, since the two conditions often occurred in the same patient (Deakin, 1996; Deakin & Graeff, 1991). They suggested that stimulation of 5-HT_2 receptors in the amygdala and frontal cortex increase sensitivity to aversive stimuli and thus experienced anxiety, but stimulation of 5-HT_{1A} receptors in the hippocampus improves resilience to stress. An opposing balance between these two types of receptors was proposed, and a distinction drawn between dorsal and median raphe function. Experimental studies investigating this theory are described (Graeff, 2004). In recent years, animal and clinical evidence has accumulated that acute serotonergic increases are associated with an inhibition of panic attacks, but increased generalised anxiety and *vice-versa* (Graeff, 2004).

Bonhomme and Esposito discussed the involvement of serotonin and dopamine in the mechanism of action of antidepressant drugs (Bonhomme & Esposito, 1998). They argued that long-term administration of tricyclics enhances the responsiveness of post-synaptic serotonin receptors to iontophoretically applied serotonin, and potentates behavioural responses to both direct and indirect dopaminergic agonists. Furthermore, they suggested that repeated administration of

selective serotonergic reuptake inhibitors and monoamine oxidase inhibitors increased serotonergic transmission by desensitising inhibitory 5-HT_{1A} somatodendritic and terminal 5-$HT_{1B/1D}$ autoreceptors, arguing that selective blockers of dopamine (DA) autoreceptors exert an antidepressant action by enhancing DA release: a similar mechanism of action was hypothesised for 5-HT_2 receptor antagonists (Bonhomme & Esposito, 1998). Whilst *acute* administration of selective serotonergic reuptake inhibitors inhibited DA neurones in the brainstem ventral tegmental area, *long term* (21 day) administration was associated with recovery from inhibition. They noted that one of the most consistent effects of long term antidepressant treatment is a down regulation of 5-HT_2 receptors, and suggested that antidepressants acting on the serotonergic system, act by resultant enhancement of the DA system. Chapter 5 provides further discussion on long versus short term antidepressant action, section 2.4.4 a discussion on the anatomy of these neurotransmitter systems.

2.4.2 Prefrontal Lobe

2.4.2.1 Orbitofrontal Cortex

Reviews of orbitofrontal cortex (OFC) function in animals are available (eg, Kringelbach & Rolls, 2004) which conclude that the OFC represents the changing and relative reward value of many unlearned (primary) reinforcers (eg, taste and somatosensory stimuli), of many different learned (secondary) reinforcers (eg, visual and olfactory stimuli), and learns and rapidly reverses associations between secondary and primary reinforcers. Specifically then, it has been proposed that the OFC implements stimulus-reinforcer associative learning, which is the type of learning involved in emotion (Kringelbach & Rolls, 2004). There is evidence that the rewarding and aversive aspects of sensory stimuli are not represented before anterior brain regions such as the OFC and amgydala (Rolls, 1999). In the case of the ventral visual pathway, processing proceeds to the stage of view invariant object recognition in the temporal lobe, before the motivational aspects of the stimuli are represented in anterior regions such as the OFC and amygdala: figure 2.4.

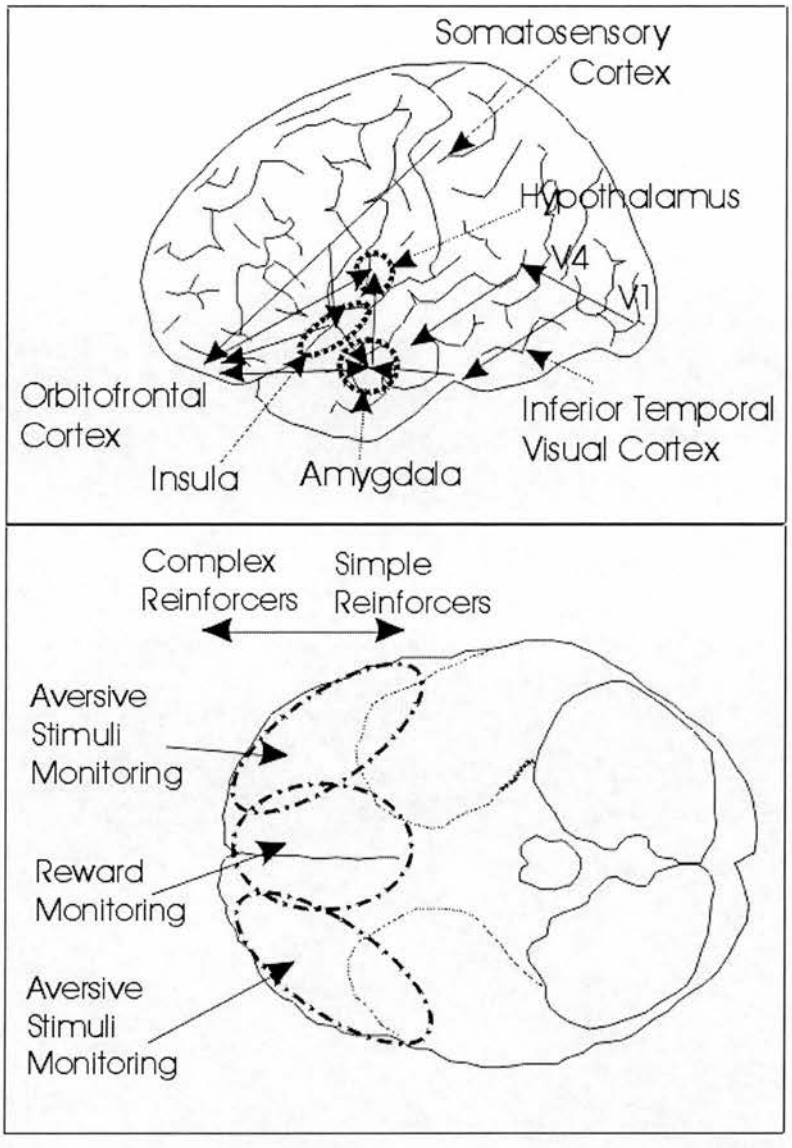


Figure 2.4. The top diagram (based on a simplification of Rolls diagram for the macaque, Rolls, 1999, figure 4.1) indicates various information processing paths leading from posterior sensory brain regions, where location and intensity of stimuli are represented, to anterior brain regions such as amygdala and orbitofrontal cortex, where motivational and emotional representation occurs. In the inferior temporal cortex, posterior to the amygdala and orbitofrontal cortex, view invariant representation of objects occurs (Rolls, 1999). The lower diagram shows a summary of orbito-frontal cortex specialisation of function according to Rolls and colleagues (Kringelbach & Rolls, 2004): the medial orbitofrontal cortex (MOFC) appears to monitor rewards, the lateral orbitofrontal cortex (LOFC) aversive stimuli.

Damage to the OFC in humans is associated with major changes in emotion, personality, behaviour and social conduct: in particular, lack of affect, socially inappropriate behaviour and irresponsibility (Kringelbach & Rolls, 2004). Such lesions in humans may severely impair the detection of some reinforcers, such as voice or face expression, responses to changing reinforcers, subjective emotion, and as a consequence, some types of decision-making (Kringelbach & Rolls, 2004). It has been proposed that these deficits occur because of impaired responding to primary reinforcers, and impairment in reversing reinforcement related associations when the contingencies change (stimulus-stimulus learning); i.e., not due to a simple inability to inhibit a previously learned motor response (Kringelbach & Rolls, 2004, p335).

A meta-analysis of 87 human imaging studies of healthy subjects reporting OFC activation has been done (Kringelbach & Rolls, 2004). It concluded that there was evidence for two distinct trends in the OFC. Firstly, there appears to be a medio-lateral segregation of function, in which the medial OFC monitors the reward value of diverse reinforcers, whereas lateral OFC activity is associated with the evaluation of punishers, associated with a possible change of behaviour. They argued that this did not show that the medial and lateral OFC have totally separate representations of reward and punishment in the human brain, since both medial and lateral regions often responded to both types of reinforcers, but in opposite ways. Secondly, there appeared to be an antero-posterior segregation of function, with more complex reinforcers (eg, monetary gain or loss) being represented more anteriorly, and simple reinforcers (eg, taste or pain), more posteriorly; i.e., a form of hierarchical representation (Kringelbach & Rolls, 2004): see figure 2.4. Overall, it was concluded that the neuroimaging studies were consistent with studies from non-human primates (Kringelbach & Rolls, 2004).

2.4.2.2 Anterior Cingulate Cortex

Detailed reviews of cingulate structure and connections are available (Devinski, Morrell & Vogt, 1995; Vogt & Gabriel, 1993). There is evidence of the anterior cingulate (AC) supporting executive behaviours and of the posterior cingulate (PC) supporting evaluative function (Vogt, Finch & Olson, 1992). Human

functional imaging studies have consistently reported PC activation in memory tasks. In contrast to the PC, the AC is strongly implicated in psychiatric disorders, and so it will be discussed further. Animal studies emphasise a general tripartite segregation of functional anatomy in the rostral to caudal direction supporting emotion, cognition and motor function. It has been argued that the emotion and cognition divisions reflect different cytoarchitecture (Devinski, Morrell & Vogt, 1995): emotion comprising Brodman Areas (BA) 25, 24 and 32 and the cognitive division predominately 24' and 32'. The boundary between the two areas is best described by a (quite precisely defined) supragenual line. The motor division of the AC comprises its most caudal area adjacent to the more superior supplementary motor area.

The distinction between the emotional division and the rest of the cingulate can be traced to trends in evolutionary development. Specifically, a phylogenetically older orbitofrontal-amygdala centred region extends throughout the emotional division of the AC, temporal polar and anterior insular regions. Additionally, a more recent hippocampal centred region extends throughout the rest of the AC and the PC (Mega & Cummings, 1997). In addition to these three broad AC divisions, other authors, on the basis of animal studies, have proposed sub-regions; e.g., a vocalization control region within the emotional division, and a nociceptive cortical region within the caudal AC (Devinski, Morrell & Vogt, 1995).

Bush and colleagues have reviewed a number of functional imaging studies reporting that the more rostral area is activated in tasks involving emotion, whereas the more caudal anterior cingulate is activated with cognitive-motor tasks (Bush, Luu & Posner, 2000): see figure 2.5. That review provides evidence that many functional imaging studies are consistent with animal and anatomical studies with regard to segregation of emotional and cognitive function. In a further link between human and animal literature, Bush and colleagues described an imaging study demonstrating significant activation of the dorsal AC (cognitive and motor region) in response to a task involving reward reduction (Bush, Vogt, Holmes, *et al*, 2002). The pattern of activation matched an earlier study in primates using electrophysiological single unit recording. In primate studies, behavioural conditioning with (e.g., electrical) stimulation of the anterior cingulate is possible (Rolls, 1999), but it is unclear

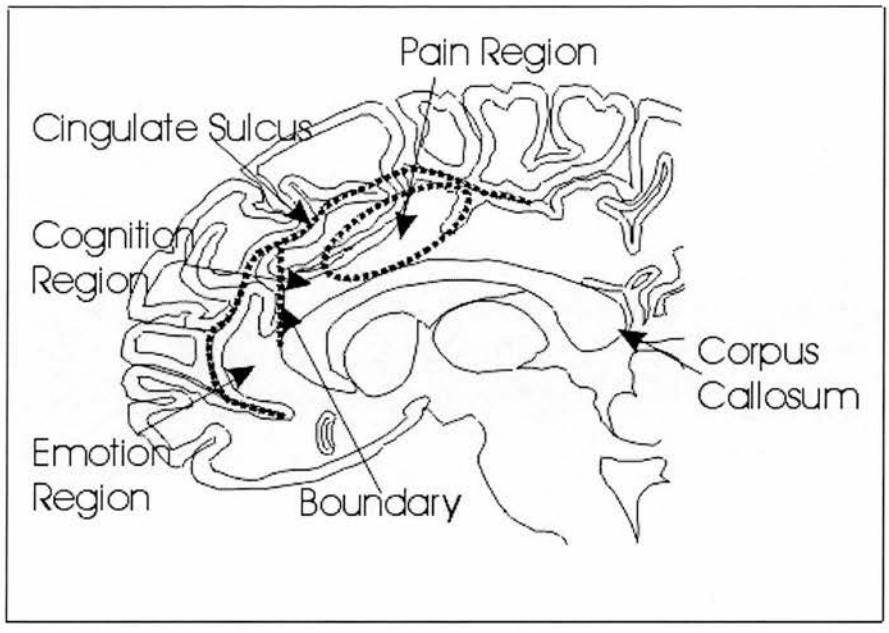


Figure 2.5 Stereotactic para-sagittal brain view of the human anterior cingulate, showing the "Boundary" between the rostral emotion and posterior cognition regions of the anterior cingulate based on an imaging meta-analysis (Bush, Luu & Posner, 2000). The ellipse encloses a sub-region reported to be active in a separate meta-analysis of imaging studies of pain (Peyron, Laurent & Garcia-Larrea, 2000): it encloses 75% of maximally active loci reported in that review. This diagram for human studies should be compared to an earlier diagram for animal studies (Devinski, Morrell & Vogt, 1995, figure 1) which appears remarkably consistent. There is evidence that the unpleasantness (affective) quality of pain is represented in the ellipse region, and not in the more rostral "emotion region" (Peyron, Laurent & Garcia-Larrea, 2000), as might be expected.

whether such conditioning can occur uniformly throughout the anterior cingulate or is localised (e.g., to just the rostral region) (Rolls, personal communication).

2.4.2.3 Dorsolateral Prefrontal Cortex

Animal and human lesion studies indicate that the dorsolateral prefrontal cortex is predominately concerned with cognitive functions, in contrast to the more posterior motor functions, and the ventromedial emotional and motivational functions (Alexander, Crutcher & DeLong, 1990; Alexander, DeLong & Strick, 1986). Damage to the dorsolateral prefrontal cortex results in a "dorsolateral syndrome" (Fuster, 1997, p172), which comprises a lack of drive and awareness, visuospatial neglect together with gaze abnormalities if the lesion extends into the frontal eye field region, preservation, and a "dysexecutive syndrome" of disruption of attention and planning. Cabeza and Nyberg have published the largest empirical review of imaging studies of diverse cognitive tasks in healthy subjects (Cabeza & Nyberg, 2000). Many cognitive tasks (but not usually emotional experiences) result in widespread dorsolateral prefrontal activation: there may be some localisation of function: eg, semantic memory retrieval appears particularly localised in and around Broca's region (Cabeza & Nyberg, 2000). Studies on depression often report underactivity, which may reflect mood related impairment of cognitive functions. Primate studies have not demonstrated behavioural conditioning using stimulation of the dorsolateral prefrontal cortex (Rolls, 1999).

2.4.2.4 Basal Ganglia Thalamocortical Loops

Figure 2.6 shows a schematic diagram of the circuitry of the basal ganglia thalamocortical loop (BGTCL) system (Alexander, Crutcher & DeLong, 1990; Alexander, DeLong & Strick, 1986). Alexander and colleagues proposed that there are at least five such partially closed re-entrant pathways linking cortex, basal ganglia and thalamus. These are named according to the cortical region: motor, oculomotor, two prefrontal (dorsolateral and lateral orbitofrontal) and limbic (medial orbitofrontal and anterior cingulate). The circuits, which are organised in parallel,

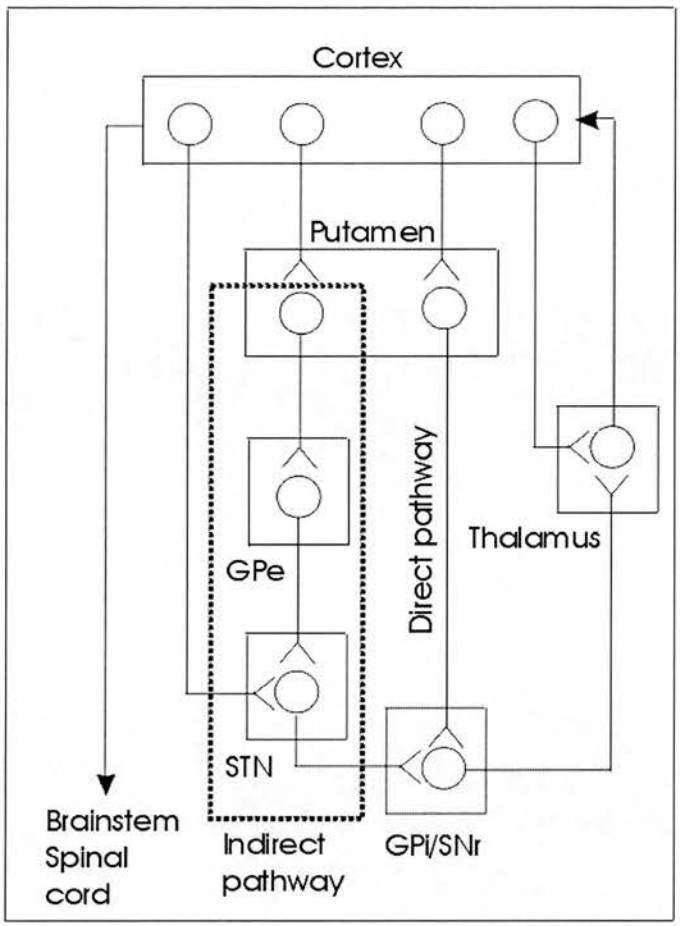


Figure 2.6 A generic basal ganglia thalamocortical loop (BGTCL). The same basic configuration is repeated within the motor, frontal eye field, dorsolateral and limbic cortical regions (Alexander & Crutcher, 1990; Alexander, Crutcher & DeLong, 1990). The diagram is simplified from the original in that the neurotransmitters are not shown, nor is the substantia nigra pars compacta, which has opposite actions on the direct and indirect pathways. Abbreviations: globus pallidus (GP) internal (i) and external (e) components, subthalamic nucleus (STN) and substantia nigra pars reticulata (SNr). The diagram should be compared with a functional representation of the same anatomical structures (chapter 4).

remain largely segregated from each other, both structurally and functionally (Alexander, Crutcher & DeLong, 1990; Alexander, DeLong & Strick, 1986). Additionally, there is evidence of a parallel architecture within the individual circuits (Alexander, Crutcher & DeLong, 1990; Alexander, DeLong & Strick, 1986).

The motor circuit has been the most extensively investigated. Other circuits are assumed to have the same general form. Whilst detailed description can be obtained elsewhere (Alexander, Crutcher & DeLong, 1990), it is worth highlighting several features. Firstly, a limited region of cortex and a similar limited region of thalamus are reciprocally connected by excitatory glutamatergic fibres. For this reason, a localised region of cortex, thalamus and connecting fibres, may be considered as a "thalamocortical unit". The internal component of the globus pallidus (GPi) and substantia nigra pars reticulata (SNr) have a high spontaneous rate of discharge and tonically inhibit the thalamus, and therefore the thalamocortical units. The output from the cortex passes via the striatum, in a predominately unidirectional corticofugal manner, to the GPi/SNr, where it acts to reduce the high spontaneous firing rate of this structure. Two main pathways from the striatum to the GPi/SNr with opposing actions have been identified: the "direct" and "indirect" pathways; the opposing actions in part being mediated by the actions of dopamine from the substantia nigra pars compacta, via different expression of dopamine D_1 and D_2 receptors on the two pathways (Arbuthnott, 1998). Input from other cortical regions occurs predominately at the cortical level. Similarly, output is predominately from the cortex in higher primates (Alexander, Crutcher & DeLong, 1990). Hypokinetic (e.g., Parkinson's disease) and hyperkinetic (e.g., Huntington's disease and hemiballismus) motor disorders reflect BGTCL dysfunction affecting the direct and indirect pathways between the striatum and GPi/SNr (DeLong, 1990).

MPTP+ treated primates are an established animal model for Parkinson's disease. A major reported finding from studies on MPTP+ treated animals is a tonic increase in GPi discharge, and enhanced phasic responses to proprioceptive stimuli and voluntary movement (Alexander & Crutcher, 1990; DeLong, 1990). This has the effect of enhanced inhibition of thalamocortical activity, and therefore of cortically initiated movement. Increased GPi activity is due to an imbalance between the direct and indirect pathways, with overactivity of the indirect pathway and underactivity of

the direct pathway (DeLong, 1990). Conversely, hyperkinetic disorders are associated with diminished GPi/SNr activity, disinhibition of thalamocortical activity, and overactivity of the direct vs. indirect pathway (DeLong, 1990). The authors of these reviews use terms such as "circuits", "gain" and "feedback", suggesting analogies with control system theory; however, they do not appear to have developed this aspect: chapter 4 therefore describes a simple control model of the BGTCLs.

Whilst the motor systems are of most interest in neurology, the prefrontal and limbic loops are of most interest in the study of psychiatric disorders (Alexander & Crutcher, 1990). In the case of the limbic loop, the ventral pallidum is not differentiated into an external and internal component, making it unclear if there are direct and indirect ventral striatal pathways (Alexander & Crutcher, 1990). Additionally, unlike other loops, the limbic loop has extensive connections with the amygdala (Alexander & Crutcher, 1990). Whilst the limbic BGTCL has been less studied than the motor system, components of the loop such as the ventral striatum, orbitofrontal and anterior cingulate cortices plus ventral tegmental area (which interdigitates with the substantia nigra, Mai, Assheuer & Paxinos, 1998) have been extensively investigated, particularly from the perspective of reinforcement learning and emotion related activity. Imaging studies of depressive illness, which implicate disordered prefrontal and limbic BGTCL function, have been discussed earlier.

2.4.3 Temporal Lobe

2.4.3.1 Hippocampus

The basic anatomy of the hippocampus consists of the same circuit being repeated over the entire septo-hippocampal length of the structure in a series of lamellae (Gray & McNaughton, 2000). There have been many models of hippocampal function: see for a review (Gray & McNaughton, 2000). One of the best established is based on experimental evidence of the hippocampus being involved in spatial cognition and memory (O'Keefe & Nadel, 1978). Another view, particularly of relevance to this thesis, is that the function of the hippocampus (or more correctly the combined septo-hippocampal system) is broader, to the extent that it has a role in emotion (Gray & McNaughton, 2000).

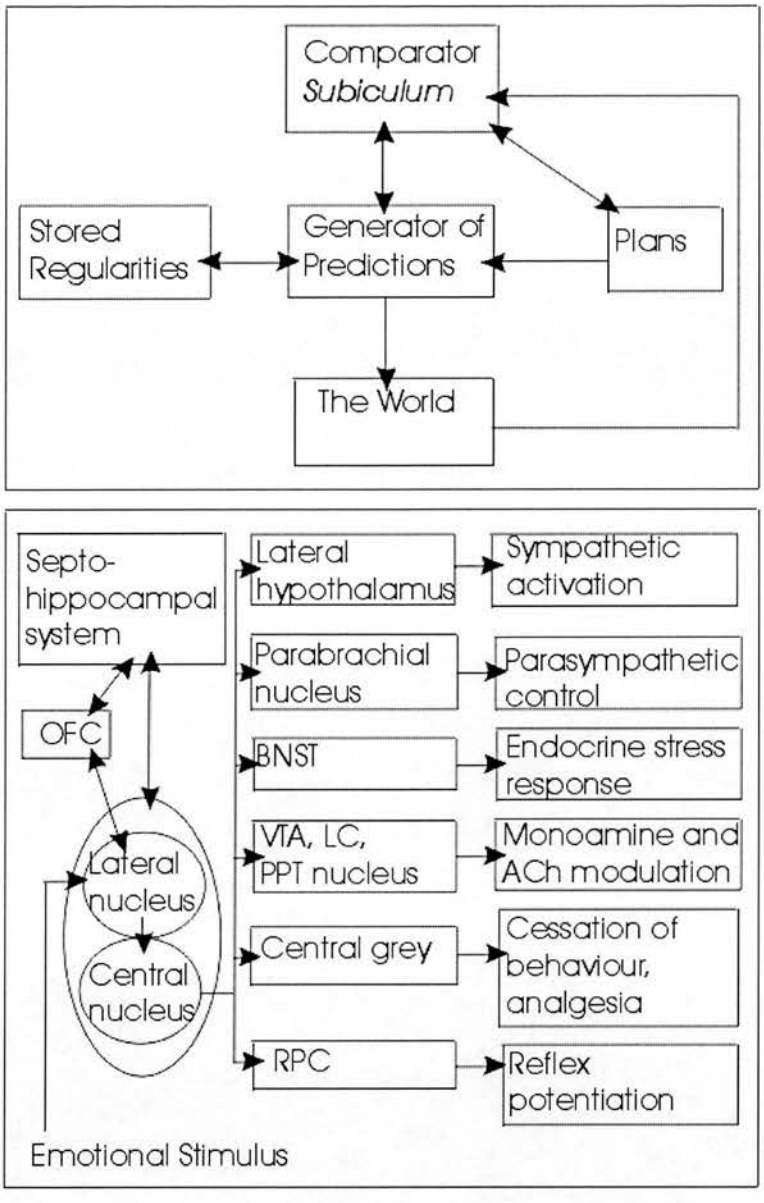


Figure 2.7 Septo-hippocampal (top) and amygdala (bottom) function. The representation of septo-hippocampal function is according to Gray and colleagues (Gray & McNaughton, 2000), who argue that the septo-hippocampal system predicts the next perceptual state, compares this with the actual perceived world, and resolves conflicting goals. The representation of amygdala function is according to various reviewers (e.g., Armony & LeDoux, 2000; Davis, Rainnie & Cassell, 1994; Gray & McNaughton, 2000). Abbreviations: bed nucleus stria terminalis (BNST), ventral tegmental area (VTA), locus coeruleus (LC), pedunculopontine tegmental nucleus (PPT), nucleus reticularis pontis caudalis (RPC), acetylcholine (ACh).

Gray and colleagues particularly focused on behavioural inhibition and anxiety, though they state that they *do not* "equate anxiety with hippocampal function", but rather view the "hippocampus as including in its information processing capacities certain operations that are crucial for the maintenance and elaboration of anxiety", and see anxiety as fundamentally resulting from "the interaction between the septo-hippocampal system and the amygdala" (Gray & McNaughton, 2000, page viii).

Gray and colleague's early model, which is based on extensive animal work, argues that the septo-hippocampal system functions as a comparator, which computes a prediction of the next likely state of the perceptual world, compares it with the actual perceived world, and generates a match or mismatch signal (Gray & McNaughton, 2000, p20): this basic operation in shown in figure 2.7. When there is a match, control of behaviour is left with the cognitive-motor brain regions (not part of the septo-hippocampal system). If the comparator detects mismatch or "threat" (predicted pain, punishment or non-reward), then the septo-hippocampal system activates the behavioural inhibition system (figure 2.2). In Gray and colleagues later model, all inputs to the septo-hippocampal system represent "goals" , the hippocampus detects conflict between currently active goals, and when conflict is detected, the hippocampus produces output which increases the valence of affectively negative stimuli and associations (Gray & McNaughton, 2000, p23). The septo-hippocampal system does not represent goals, but resolves competing goals through its actions on other brain regions (e.g., amygdala, hypothalamus, premotor regions). Presumably, input goals might come from regions such as the amygdala, anterior cingulate and additionally the orbitofrontal cortex.

2.4.3.2 Amygdala

Aggleton's text on amygdala function is one of the most comprehensive currently available (Aggleton, 2000). Extensive studies on fear conditioning in animals have provided clear evidence of the amygdala being directly involved with anxiety (Davis, Hitchcock & Rosen, 1987; Davis, Rainnie & Cassell, 1994; LeDoux, 1998a) and probably all other emotions (Cardinal, Parkinson, Hall, *et al*, 2002). A diagram of the neural circuits involved in fear conditioning is shown in figure 2.7. The lateral nucleus of the amygdala receives input from sensory areas including the

thalamus, as well as the hippocampal formation (figure 2.7) and cortical association regions, and via intra-amygdala processing, the information reaches the central nucleus, which controls the expression of different components of the affective response. There appears to be a bias in the literature suggesting that the amygdala is particularly involved in the mediation of aversive emotions (e.g., anxiety), in contrast to positive emotions. Controversy exists as to whether such bias is a genuine reflection of the physiology, or an artefact of the relative ease with which aversive emotions are invoked in laboratory settings. The amygdala has been implicated as having abnormal function and perhaps structure in depressive illness. A detailed review of the role of the amygdala, ventral striatum and prefrontal cortex in relation to normal emotion, motivation and associative learning can be found elsewhere (Cardinal, Parkinson, Hall, *et al*, 2002).

2.4.4 Brainstem

Antidepressants have actions on all the main classical neuromodulator systems: in particular serotonin and noradrenaline, but also dopamine and acetylcholine. The nuclei of cells, which synthesise these chemicals, are located in the brainstem. Figure 2.8 shows a sagittal view of the brain together with a transverse section through the brainstem at the level of the midbrain-pons junction. The raphe nuclei cells synthesise serotonin, locus coeruleus noradrenaline, and the ventral tegmental neurones dopamine, for other brain structures, including the limbic region. These nuclei are elongated along the axis of the brainstem and extend well into the pons. The bilateral substantia nigra cell bodies project mostly to basal ganglia motor nuclei although these overlap with the VTA cells. In animals, the periaqueductal grey (PAG) matter in the brainstem has been reported to be functionally segregated along the axis of the brainstem with regard to behaviour (Bandler & Keay, 1996). The lateral PAG is associated with active defensive behaviour (fight and flight), hypertension, tachycardia and non-opioid analgesia.

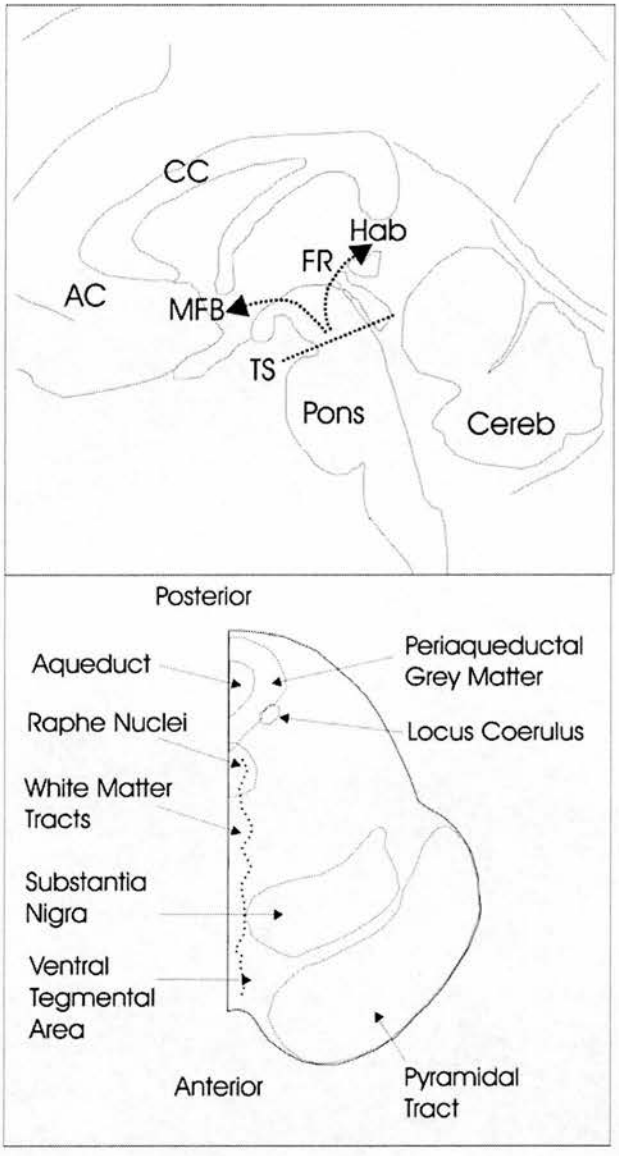


Fig 2.8 Sagittal and transverse view of the brainstem. TS indicates the location of the transverse slice through the inferior midbrain. Abbreviations: anterior cingulate (AC), corpus callosum (CC), medial forebrain bundle (MFB), fasciculus retroflexus (FR), habenular nucleus (Hab), cerebellum (Cereb). The "White Matter Tracts" include the MFB and FR. Chapter 6 describes an investigation into a reported abnormality of these tracts in depressive illness.

Reviews of the *normal physiological role* of these monoamines and ACh have been published. Dopamine (DA) is strongly implicated in motor and behavioural reward mechanisms, noradrenaline (NA) in non-specific arousal and selective attention (Robbins & Everitt, 1995). In the case of serotonin (5HT) however the role is less clear. As discussed above, one theory specifically implicates serotonin in the mediation of defensive behaviour and anxiety (Graeff, 1993) and is included as such in Gray's behavioural inhibition theory of septo-hippocampal function (Gray & McNaughton, 2000). However, in general it is difficult to ascribe a single unitary role for serotonin and there is significant evidence of serotonin acting in opposition to the other major neuromodulators (Robbins & Everitt, 1995). This emphasises the fact that all neurotransmitters do not act in isolation but instead interact. All the classical neuromodulator systems (DA, NA, 5HT, ACh) are components of the more general reticular activating system (Robbins & Everitt, 1995).

The fibre tracts connecting these nuclei to the rest of the brain pass mostly along the midline (Lang, 1993). One tract, the medial forebrain bundle, then runs anteriorly through the lateral hypothalamus and ventral striatum to more distant prefrontal cortical regions implicated as structurally and functionally abnormal in imaging studies of psychiatric disorders. Another, the fasciculus retroflexus, passes posteriorly to the habenular nucleus in the dorsal diencephalon. It has long been known that direct electrical stimulation of the medial forebrain bundle in animals is rewarding (i.e. animals will work to repeat the stimulation) (Olds & Olds, 1964), and more recently, that this is specifically due to stimulation of the dopaminergic fibres (Rolls, 1999). As noted earlier, Rolls argues that stimulation (whether by electricity or drugs) of this tract (and projection sites such as orbitofrontal cortex) is rewarding *because* it comprises part of the system normally controlling appetitive behaviour (Rolls, 1999). In animals, the fasciculus retroflexus degenerates in response to sustained high levels of various drugs, including stimulants and nicotine, which may mimic the effects of human binge use (Ellison, 2002). This tract is involved in the regulation of brainstem neurotransmitter activity and has been reported as exhibiting abnormal connectivity in a tryptophan depletion study of depressive illness relapse (Morris, Smith, Cowen, *et al*, 1999). The "white matter tracts" shown in figure 2.8

have been reported to be structurally abnormal in various studies of unipolar depressive illness (Becker, Berg, Lesch, *et al*, 2001): chapter 6 describes an attempt to replicate these findings.

2.4.5 Quantitative Neural Network Models

The experimental study of functional segregation addresses the question of which functions are performed by different brain regions. In contrast, computational models are useful in addressing the issue of *how* they may be performed, and which aspects of brain function might be important for achieving the function of interest (Rolls & Treves, 2001): the latter is the focus of chapter 4, a brief introduction is provided here. Two subsections are provided, one on functional segregation itself, the other on reinforcement (emotional) learning. The equations which follow are included since this work is necessarily mathematical. As in other areas of science such as physics, the equations and not an approximate description, are the theory.

It is important to discuss the term "model", since this is used in a number of places in this thesis. In common usage the term refers to a *representation* of an object. Usually this is of a selected feature and not all features. The term "internal model" is popular in the engineering sciences and denotes a set of equations that describe the temporal development of a real world process (Garcia, Prett & Morari, 1989). Studies in cognition, brain theory and motor control suggest that humans and animals associate events (stimuli, reinforcers or behavioural responses) with other events, and use these associations to form novel *associative chains* (see for references and discussion, Suri, 2001). Since the quantitative internal model approach computes event-specific prediction signals and uses these signals to simulate hypothetical future experience, the internal model approach reproduces the formation of novel associative chains (Suri, 2001; Sutton & Barto, 1998).

This is of particular relevance to cognitive science which has the central hypothesis that information about the world is interpreted and experienced in the form of models comprising mental *representations* distinct from objects. Yates, in a comprehensive review of these concepts in psychology, concluded that the brain generates models of the world capable of simulating future events, anticipating present events, and thereby formulating appropriate actions (Yates, 1985). The

origin of such psychological concepts dates back at least to Kant (who introduced the term "schema") and Helmholtz ("categories of experience"), and possibly to the much earlier Platonic doctrine of "ideal types". Despite this very long history, the aim of understanding the neural substrate of psychological models or representation has only recently been recognised as being of fundamental importance to theoretical neuroscience (Eliasmith & Anderson, 2002). Central to this work are concepts of prediction, control and adaptation.

Consider a neuron (i) which receives axonal synaptic contact from another neuron (j). Neuron i makes a summation of its dendritic inputs (synaptic currents, r'_j) defined as

$$h_i = \sum_j r'_j w_{ij} \qquad 2.1$$

where w_{ij} is the strength of the synaptic connection (Rolls & Treves, 2001). This produces activation at the cell body, resulting in a firing rate

$$r_i = f(h_i) \qquad 2.2$$

where $f(.)$ is a function representing firing above a given threshold. A network of such neurones can produce a given output, for a particular input, if w_{ij} is appropriately learned. Hebb suggested a simple learning rule (Hebb, 1949)

$$\delta w_{ij} = k r_i r'_j \qquad 2.3$$

where δw_{ij} is the change in w_{ij} which results from the "simultaneous" (probably within 100-150 ms in the brain, Rolls & Treves, 2001) presence of presynaptic firing r'_j and postsynaptic firing r_i, and k is a constant determining how much the synapses alter on any one occasion. Hebb's rule reflects the features of long term potentiation (LTP) in the brain: many other synaptic modifications also occur; e.g., long term depression (LTD) (Rolls & Treves, 2001). From a computational perspective, use of

equation 2.3 is impractical due to problems with interference between input patterns, and a modified Hebb rule is often used (Rolls, 1999, p313)

$$\delta w_{ij} = k r_i \left(r_j' - x \right) \qquad 2.4$$

where x is a constant, approximately equal to the long term average of r'. Equation 2.4 combines the features of both LTP and LTD (Rolls & Treves, 2001). Further general discussion on quantitative neural network models and synaptic plasticity in relation to psychiatric disorders can be found elsewhere (Jeffery & Reid, 1997). Siegle has additionally described a detailed neural network model of attention biases in depressive illness (Siegle, 1999). Functional segregation and reinforcement learning is of particular relevance to this thesis and so these topics will now be briefly mentioned.

2.4.5.1 Functional Segregation

Artificial neural networks can be classified as involving unsupervised or supervised learning. Considering unsupervised learning, figure 2.9 shows a Kohonen self-organising map (SOM). This comprises a single flat grid representing a layer of neurones forming the cortical surface. Additionally, there is a node layer which is both the input and output from the grid: it might represent thalamic connections. Details of the various algorithms, which allow such networks to develop topological maps (in which the most important similarity relationships between the input signals are represented by the spatial relationships of the neurones), can be found elsewhere (e.g., Beale, 1990). The crucial aspect of the cortical grid allowing self organisation is the assumption of short range excitation and longer range inhibition: a "Mexican hat" function (figure 2.9). Short range excitatory connections, which diminish over about 1 mm, and longer range inhibitory connections, are commonly present in the brain, including the cortex (Rolls & Treves, 2001, p67). Additionally, the learning rule which is required for such algorithms, is a modified form of Hebb rule, so is also biologically plausible (Rolls & Treves, 2001). More complex models which use this approach have been described, e.g., to model the tonotopic organisation of the auditory cortex (Mercado, Myers & Gluck, 2001) and used to describe the

development of visual topographic receptive fields, explaining why lateral connection patterns closely follow receptive field properties such as ocular dominance (Sirosh & Miikkulainen, 1997). Topographical map formation results in the different features of objects being represented in a two dimensional plane. In the case of simple objects, such representation results in a systematic variation in the represented features across the map. However, in the case of complex objects, which can only be represented in a high number of dimensions, dimensional reduction (c.f., multidimensional scaling) to a two dimensional cortical plane results in fractures or discontinuities in the map. This has been suggested as occurring, for example, in the inferior temporal lobe (Rolls & Treves, 2001).

2.4.5.2 Reinforcement Learning

Following invariant pattern representation, reinforcement learning occurs in anterior brain regions such as the amygdala and orbitofrontal cortex. Pattern association between a primary reinforcer (e.g. taste of food) and a potential secondary reinforcer (e.g., sight of food) has been suggested (Rolls & Treves, 2001, p150) to involve the network shown in figure 2.10. Temporal cortical visual signals, which do not exhibit habituation, reach the amygdala where habituation does occur, unless there is co-occurrence of the visual stimuli with a primary reinforcer (rewarding or aversive) (Rolls, 1999). Another mechanism, and of particular relevance in this thesis, involves supervised reinforcement learning (Rolls, 1999; Rolls & Treves, 2001). In this case a global error (reinforcement) signal is provided for the network which specifies the magnitude and direction of the error in the network output for given input. The error signal is used to correct the synaptic weightings such that the output error diminishes with time; e.g., Sutton and Barto's "associative reward-penalty" network is shown in figure 2.10 (see discussion and references in, Rolls, 1999, p318). The synaptic weights are changed according to learning rate terms, which are higher when positive reinforcement is achieved. The temporal difference (TD) algorithm (chapter 4) allows learning to occur when the reinforcement is delayed or received over many time steps.

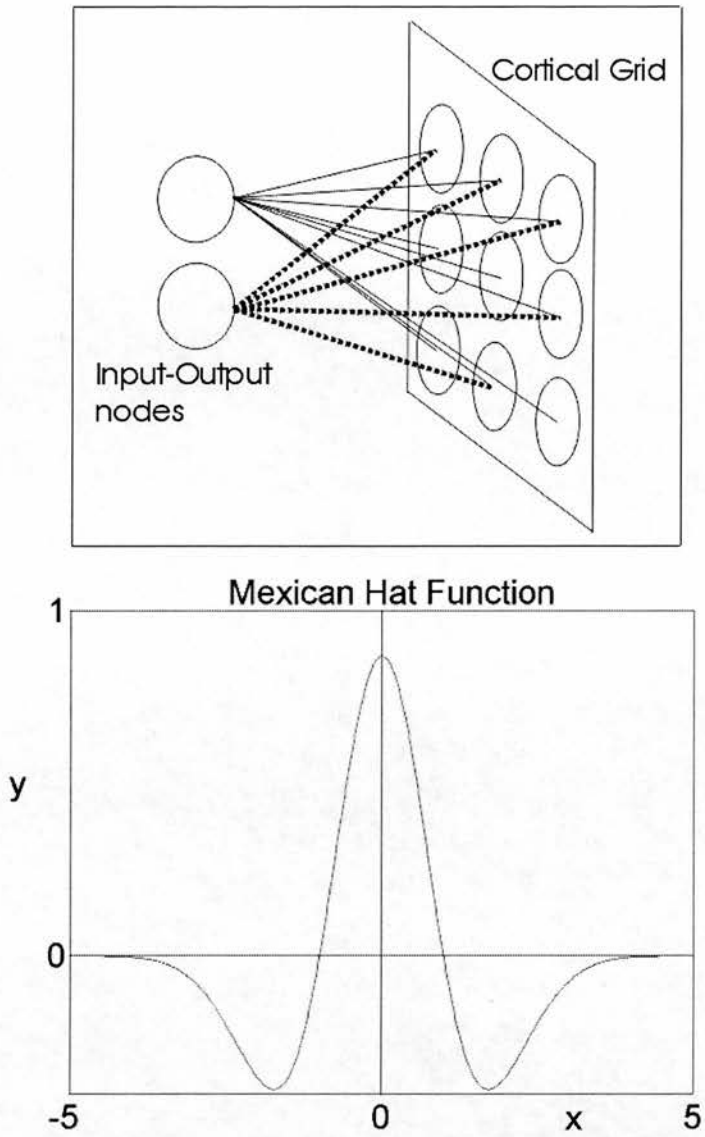


Figure 2.9 The top diagram shows a Kohonen self-organising map (SOM). Signals are input via the nodes, processing occurs in the 2 dimensional grid, then output occurs via the nodes. The cortical grid develops a segregated representation of input activity in a variety of situations, which all involve an assumption of a Mexican Hat (shown below) type interaction, comprising center excitation and surround inhibition, between neighbouring neurones in the cortical grid (calculated using the Wavelet Toolbox, Matlab).

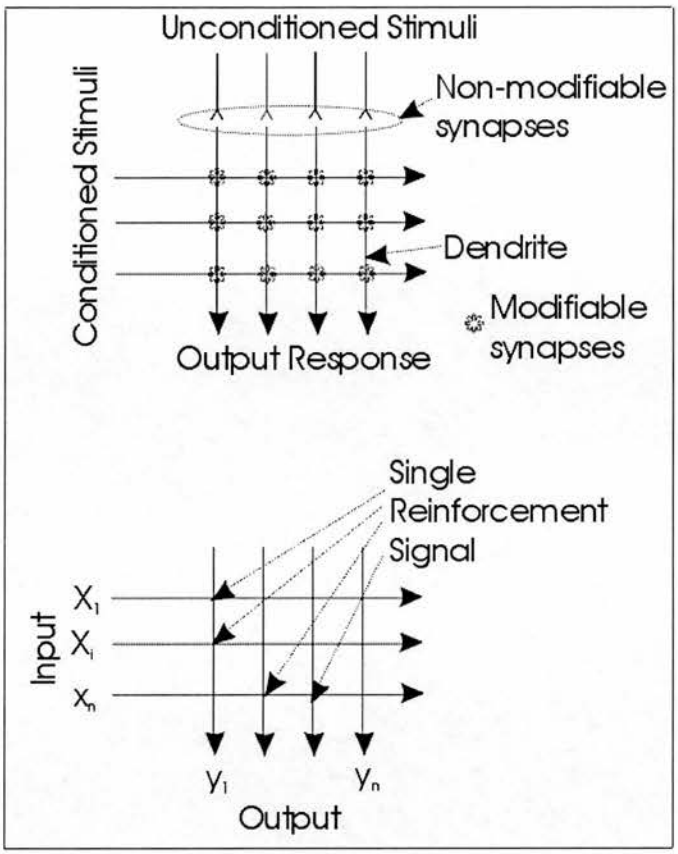


Figure 2.10 Distributed processing models of associative learning in the human brain. The top diagram shows a network for pattern association between a primary reinforcer and a potential secondary reinforcer: this may occur in the amygdala. The modifiable synapses change according to a local Hebbian law: there is no single global reinforcement signal. The lower diagram shows Sutton and Barto's associative learning network which has a single global error or reinforcement teaching signal which modifies the synaptic weightings. There is extensive experimental evidence of the monoamine systems, and many other brain structures, representing such error or reinforcement signals (see chapter 4). Chapter 5 describes an investigation into a hypothesised abnormality of a global error signal in depressive illness.

There is extensive experimental evidence from work in animals, and more recently replicated imaging work in humans, for various brain structures, including the monoamine systems, exhibiting such global error signals (chapter 4).

2.4.6 Functional Segregation and Subjective Experience

Imaging studies of depressive illness seek localised functional and structural abnormalities, hypothesising a relationship with subjective experience (symptoms). Consequently, it is necessary to consider subjective or conscious experience in relation to brain functional segregation. The question of where and how consciousness arises has been a topic of interest to philosophers for millennia: there is consequently a considerable literature. Zeman however provides an introduction (Zeman, 2002). The focus here will be on one (scientific) theory, which although confined to the visual system, nonetheless may be applicable to other brain regions of more interest to psychiatric research.

Zeki and Bartels note that anatomical and physiological studies indicate that the primate visual system consists of many distributed processing systems acting in parallel. They emphasise psychophysiological studies and interpret them as demonstrating that neural activity in each of the parallel systems reaches its perceptual end-point at a slightly different time, leading to "perceptual asynchrony in vision". They argue that this, combined with clinical and imaging evidence, suggests that the processing systems are also perceptual systems, with each system acting in a semi-autonomous manner (Bartels & Zeki, 1998). Additionally, they argue that such activity can have a conscious correlate ("microconsciousness") without necessarily involving activity in other systems, and conclude that visual consciousness is itself modular, reflecting the basic modular organisation (functional specialisation or segregation) of the brain. Zeki and Bartels cite a substantial amount of experimental evidence as consistent with their theory. Having argued for a fundamentally "fractionated" basis of visual consciousness, they then consider the issue of integration of microconsciousness to form a perceptual unity (Zeki, 2001).

Zeki and Bartels' theory is important because it explicitly argues that segregation of function in the brain may be associated with a corresponding conscious correlate. Presumably then, functionally segregated regions of the

prefrontal and anterior temporal lobes could give rise to other components of consciousness; eg, emotion. Deficits resulting from lesions in humans might be interpreted as consistent with this conjecture. For example, Goodwin comments: "Are we in a position to say whether a lesion in the inferior frontal areas renders people incapable of experiencing emotion ? The answer is, probably, yes... the failure of [such a person's] on-line monitoring of interoceptive feelings seems to deprive them of 'gut feelings' that are essential for normal cognition" (Goodwin, 1998, p415). This deficit might be interpreted as a selective absence of emotion related subjective experiences which are components of an overall normal consciousness.

In conclusion, understanding the dysfunctional neural mechanisms of abnormal subjective experience (e.g., abnormal emotions and hallucinations) is of central importance to understanding psychiatric disorders. Consequently, approaches to studying conscious experience and the function of posterior brain regions (e.g., vision) may also be relevant to the study of psychiatric symptoms and anterior brain function (e.g., emotion and mood).

2.4.7 Reinforcer Representation and Emotion

There have been many different theories of emotion: see elsewhere for reviews (Rolls, 1999). However, it is noticeable that the brain regions reported as representing the rewarding and aversive aspects of stimuli in animals, and which provide the neural substrate for stimulus-response learning, appear to correspond to the brain regions reported as active when humans experience emotion (chapter 3).

Consequently, it is of interest that emotions have been defined as states produced by instrumental reinforcing stimuli (Rolls, 1999). This view is consistent with previous theories (Gray, 1981; Millenson, 1967; Weizcrantz, 1968). Instrumental reinforcers are stimuli which if their occurrence, termination or omission is made contingent upon making a response, the probability of the emission of that response is altered (Rolls, 1999). This definition does not require that all such stimuli need to be formally shown to be instrumental reinforcers (Rolls, 1999). Some stimuli are unlearned (primary) and others are learned by association (secondary).

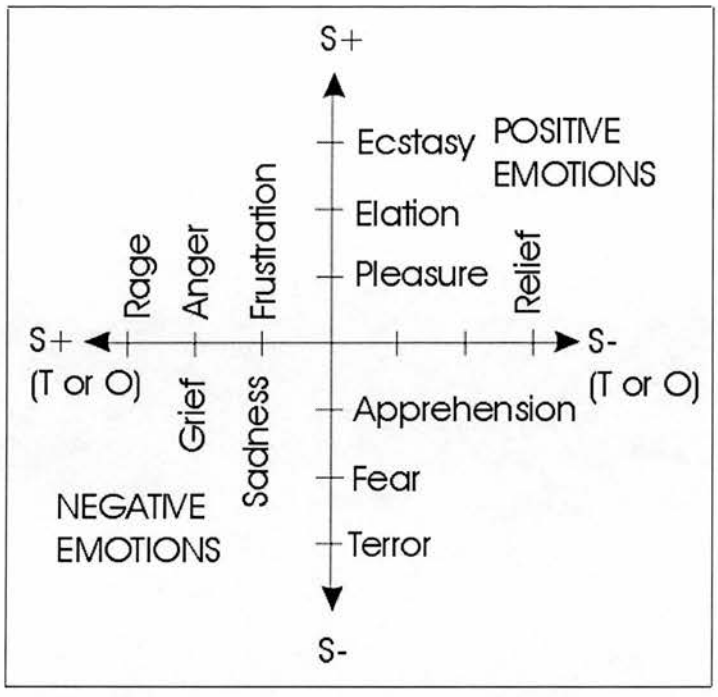


Figure 2.11 Rolls' theory which links normal human emotion to positive (+) or negative (-) reinforcers (S), and their termination (T) or omission (O) (Rolls, 1999). Intensity of emotion is represented by distance from the intersection of the axes. This diagram categorises emotions as either positive or negative and with varying intensity, reflecting the underlying reinforcers. The study described in chapter 5 linked the presentation of unpredicted reinforcers (horizontal axis of the above diagram) with emotional response via this theory.

A positive reinforcer (reward) increases the probability of emission of the response on which it is contingent; conversely a negative reinforcer (punisher) reduces the probability of such a response. For example, an emotion (fear) results from learning a stimulus (tone) – reinforcement (electrical shock) association. Two advantages of this approach are that it results in an operational definition of an emotion, and permits classification of emotions in terms of different reinforcement contingencies (Rolls, 1999). The main features of this theory are shown in figure 2.11. Two axes are present with positive (+), negative (-), omission of expected (O), and termination of ongoing (T) reinforcers represented. Intensity of emotion increases with distance from the axes intersection. Rolls argues that a very wide variety of emotions can be described and classified according to many factors including: whether the reinforcer is positive or negative, the reinforcement contingency, the combinations of reinforcers in a given situation (e.g., simultaneous reward and punishment producing conflict and guilt) and responses which are possible in a given environment (e.g., anger if an avoidable aversive stimulus is presented, fear if it is not avoidable) (Rolls, 1999). There is a substantial opportunity for cognitive processing affecting emotions, in that cognitive processing will often be required to categorise a stimulus as rewarding or aversive (Rolls, 1999).

2.5 SUMMARY AND HYPOTHESES

Regions which are repeatedly reported as having abnormal function and sometimes structure in depressive illness, such as the orbitofrontal and anterior cingulate cortices, amygdala and hippocampus, and the components of the limbic BGTCL, are the substrate in animals for associative learning and the representation of the rewarding or aversive aspects of stimuli. Normal human emotional experience has been linked to the presentation or omission of reinforcers. When this is considered in the context of the clinical features of depressive illness (e.g., anhedonia and anxiety), it suggests that depressive illness may comprise a disorder of associative learning. Chapter 5 describes a test of this hypothesis. Furthermore, the monoamine systems, which are an important component of the substrate for associative learning, are partly located in the brainstem. The observation that antidepressants directly affect this system provides a further link between associative

learning and depressive illness. Previous studies reported a subtle anatomical abnormality in the brainstem of a high proportion of patients, suggesting a possible direct cause of depressive illness symptoms. Chapter 6 describes an attempt at replication.

A useful clinical distinction is made between cognitive and emotional processes, and there is evidence that such processes may be segregated in the anterior cingulate. However, it is not known whether such segregation exists more widely in the prefrontal lobe. Some imaging studies support the hypothesis that the ventromedial prefrontal region is associated with emotional experience, and the dorsolateral cortex with cognitive processing; however, there is a great deal of inconsistency between individual reports. It is therefore interesting to try to establish whether the ventromedial prefrontal cortex does tend to be activated in association with healthy human emotional experience, in contrast to other prefrontal regions. In part, this is because it could provide additional support for the hypothesis that the same brain regions, which support associative learning in animals, correspond to those associated with emotional experience in humans (Rolls, personal communication). Such an investigation is described in the next chapter.

Finally, Rolls theory of emotion is clearly an approximation, since it does not take account of evidence for the emotional response changing with repeated presentation of a given reinforcer (Solomon, 1980a). Future descriptions will have to take account of such change, and in so doing, it may be useful to quantify the affective response (eg, Lang, Bradley & Cuthbert, 1998)

CHAPTER 3

SEGREGATION OF COGNITIVE AND EMOTIONAL FUNCTION IN THE PREFRONTAL CORTEX: A STEREOTACTIC META-ANALYSIS

3.1 INTRODUCTION

Reviews of imaging studies of patients with major depressive disorder and schizophrenia consistently implicate the prefrontal cortex (Drevets, 1999; Ebmeier & Kronhaus, 2002; Lawrie & Abukmeil, 1998; Wright, Rabe-Hesketh, Woodruff, *et al*, 2000). For example, focal structural and functional abnormalities in the subgenual anterior cingulate have been reported in patients with depression (Drevets, Price, Simpson, *et al*, 1997) and in patients with schizophrenia (Job, Whalley, McConnell, *et al*, 2002). The orbitofrontal cortex has similarly been reported as abnormal in both disorders (Drevets, 2000b; Sanfilipo, Lafargue, Rusinek, *et al*, 2000).

The interpretation of these findings is however problematic given the limited understanding of normal function. Prefrontal function may be segregated but the definition of the spatial limits of such segregation is not well established. A significant contribution has been made by Bush and colleagues who reviewed human functional imaging studies and demonstrated segregation of emotional and cognitive function in the anterior cingulate (AC) (Bush, Luu & Posner, 2000). The boundary between these two regions is supragenual and appears consistent with differences obtained from studies of cytoarchitecture, and the effects of lesions and electrical stimulation (Vogt, Finch & Olson, 1992; Vogt, Nimchinsky, Vogt, *et al*, 1995).

Studies of phylogenetic isocortical development (Mega & Cummings, 1997, figure 10.2) identify two trends: a rostral paleocortex uniting the orbitofrontal, insula and temporal polar regions which extends into the rostral AC (rAC), and a more caudal hippocampal archicortex uniting the parahippocampal and entorhinal regions, and extending into the posterior cingulate, caudal AC (cAC) and rAC. The rostral region is phylogenetically older and processes the internal affective state of the organism; the caudal archicortical region is concerned with external "evaluative" processing of stimuli (Mega & Cummings, 1997).

The rAC and orbitofrontal cortex (OFC) project to structures such as the hypothalamus and periaqueductal grey matter (Ongur & Price, 2000). In addition, the rAC and OFC are part of a thalamocortical basal-ganglia re-entrant "limbic" loop which includes the ventral striatum and pallidum and connects extensively with the amygdala (Alexander, Crutcher & DeLong, 1990). The medial forebrain bundle, which includes axonal projections from brainstem monoaminergic nuclei, projects to the AC and OFC. All these structures have long been associated with emotion (McLean, 1949; Olds & Olds, 1964; Rolls, 1999); e.g., electrical stimulation tends to evoke emotional, behavioural and autonomic changes, whilst damage to these regions is associated with emotional, social and behavioural deficits (Fuster, 1997).

In the medial prefrontal cortex, the cAC has few connections with the amygdala, periaqueductal grey matter and hypothalamus, and electrical stimulation and lesion studies do not link it to emotional function (Devinski, Morrell & Vogt, 1995). Instead, it is implicated in diverse cognitive functions with its most posterior region contributing to skeleto-motor control. Detailed maps of body representation have been described in this region of primates (Dum & Strick, 1993) and electrical stimulation in humans evokes integrated motor actions (Talairach, Bancaud, Geier, *et al*, 1973). The division between the cognitive and emotional regions of the medial prefrontal cortex is best defined by a line drawn just caudal to the border of Brodman's Area (BA) 32 to include most of areas 24, 25 and 33 (Bush, Luu & Posner, 2000; Devinski, Morrell & Vogt, 1995, p287).

Damage to the dorsolateral prefrontal cortex causes cognitive deficits, whereas damage to the OFC is associated with a different pattern of emotional and social deficits (Fuster, 1997). However, it is unclear where on the lateral cortical surface the division between emotional and cognitive function should best be drawn. For example, Alexander provides evidence that the lateral OFC is distinct from the medial "limbic" orbitofrontal and dorsolateral cortices (Alexander, Crutcher & DeLong, 1990), yet some parts of the lateral OFC are strongly connected to the medial region (Price, 1999). A meta-analysis of human imaging studies may help to clarify this issue.

Various meta-analyses of functional activation studies already exist (Bush, Luu & Posner, 2000; Cabeza & Nyberg, 2000; Duncan & Owen, 2000; Phan,

Wagner, Taylor, *et al*, 2002; Turkeltaub, Eden, Jones, *et al*, 2002). In all cases, a variable number of reported activation loci from each study is included. Unfortunately, this does not take into account differences in study power and other potential sources of bias (Lawrie, McIntosh & Rao, 2000). Bias can arise in observational meta-analyses in many ways and methods to reduce such unwanted effects from e.g. publication bias, are now routinely used (Stroup, Berlin, Morton, *et al*, 2000).

The object of this study was to further investigate segregation of cognitive and emotional function in the medial and lateral prefrontal cortex. The studies considered here comprise those already identified in the two largest reviews of cognitive tasks and emotion induction yet published: Cabeza discussed 275 imaging studies involving a wide range of cognitive tasks (Cabeza & Nyberg, 2000) and Phan discussed 55 emotion induction tasks (Phan, Wagner, Taylor, *et al*, 2002). Both reviews are comparable in that they include studies published over essentially the same decade, consider only findings in healthy volunteers, and include only PET and fMRI studies. The definitions of cognitive tasks and emotion induction used here follow the earlier publications. Cabeza defined cognitive tasks as including attention, perception, imagery, language, working memory, semantic memory, episodic memory, priming and procedural memory. Phan defined emotion induction as including happiness, fear, anger, sadness, disgust, positive (various pleasant emotions) and negative (various unpleasant emotions).

The aim of this meta-analysis was to extract information from each of the 330 studies in a standardised reproducible manner in order to minimise bias. At most one medial and one lateral activation loci were extracted from each study. The effects of variation in study power on activation patterns were investigated. The spatial distributions of reported activations for cognitive tasks and emotion induction were hypothesised to differ. In the case of the medial cortex, the segregation reported previously by Bush was hypothesised to extend beyond the boundaries of the anterior cingulate. In the case of the lateral cortex, the orbitomedial region was hypothesised to be most active in emotion induction tasks with the dorsolateral region being most often active in cognitive tasks.

No prediction was made regarding the hemisphere of activation. Whilst there is good evidence for lateralisation of cognitive functions such as speech and visuospatial processing, the evidence for lateralisation of other functions is not so good (Cabeza & Nyberg, 2000) and there is only limited evidence for emotional processes being lateralised. Therefore, since an examination of lateralisation would reduce study power, we chose to omit it.

If justified by rejection of corresponding null hypotheses, best estimates of most likely reported activation loci were planned, together with estimates of uncertainty. This allows derivation of probability maps for activations and assists in the interpretation of reported focal abnormalities in patient groups. It also allows subsequent testing of *a priori* defined regions of interest in future patient studies. Such testing has various advantages including a reduction in the need for multiple testing of voxels with consequent increase in study power, and easier interpretation of results.

3.2 MATERIALS AND METHODS

3.2.1 Selection of Activation Loci

Although the aim of this study was to extract activation loci data from the two previous reviews, additional searches were done to determine whether it was possible to identify a significant number of missed studies. Searches were done using standard databases (MEDLINE, EMBASE, PsychINFO) using very general terms to maximise the number of identified studies. Abstracts were then inspected to determine whether the studies were already identified, and if not, whether they might be suitable. The reference sections of the already included studies were similarly inspected, and this took much longer than the on-line searches. To determine whether a study was suitable, the same criteria were used as in the previous reviews (Cabeza & Nyberg, 2000; Phan, Wagner, Taylor, *et al*, 2002) . Studies were suitable only if they investigated unmedicated healthy adults without neurological or psychiatric disorder, investigated higher order brain function (as defined in the previous reviews) and reported results using a voxel based method. Using this method, it was not possible to identify a significant number of missed studies.

In areas implicated in emotional processing, cerebral blood flow tends to increase during emotion related tasks, but decrease during some attentionally demanding cognitive tasks. Conversely, in regions implicated in cognitive processing, blood flow tends to increase during attentionally demanding cognitive tasks, but decrease during some normal and abnormal emotional states (Bush, Luu & Posner, 2000; Drevets & Raichle, 1998). Therefore, to simplify this analysis, only activations were considered. Both authors separately inspected each individual publication cited by Cabeza and Phan to select reported activation coordinates according to the criteria given below. Any disagreements were resolved by discussion.

The prefrontal cortex was defined conventionally (Fuster, 1997) as excluding primary motor, premotor and frontal eye field areas: specifically BA 6, 7 and 8. Loci were assumed to be reported in MNI stereotactic space, since image processing templates conforming to such space are most commonly used. Inclusion or exclusion of those points which occurred near the boundary of these regions was determined by converting these points to Talairach space using a method described by Brett (http://www.mrc-cbu.cam.ac.uk/Imaging/) and the corresponding BA determined using a database (http://ric.uthscsa.edu/projects/talairachdaemon.html).

Only studies reporting results obtained using voxel based analysis methods were included. A single maximally significant activation locus, which was not limited to the anterior cingulate, was determined for the medial prefrontal cortex. Significance was determined by reported z score or t statistic. Occasionally these measures were absent for individual loci and in this case the largest reported activation volume was chosen. If the activation loci from a study were reported only in terms of BAs (without coordinates) the study was omitted. All included loci had to represent a significant activation at a conventional level (e.g. at least at $p<0.001$ for fMRI studies). If it was not possible to determine the most significant activation locus within this region, the study was omitted. The hemisphere of activation was disregarded. The contrast for the reported activation could be either a cognitive task or rest for emotion induction, another cognitive task or rest for a cognitive task. The lateral prefrontal cortex was examined separately using identical criteria.

Thus for each study at most two activation loci were recorded. Additionally, the number of subjects used to determine the activation locus was recorded, as was the type of study (PET or fMRI). Finally, if it appeared that a very similar study with identical subjects had already been included, only the largest such study was included. This was done to ensure the independence of each activation locus.

3.2.2 Initial Data Transformations

Summary estimates of the centres of activation for cognitive and emotional tasks in the medial and lateral cortices were planned. However, the anterior cingulate curves through 180° in the sagittal plane and the lateral cortical surface curves through a similar angle in the coronal plane. Therefore, the activation loci were transformed (unwarped) into different geometrical spaces before hypothesis testing and summary statistic determination. Figure 3.1 shows the method used to unwarp the medial prefrontal cortex loci. A series of points were determined which lie on the exterior surface of the corpus callosum obtained from a high resolution T_1 weighted MRI scan of a normal subject with the anatomical scaling corresponding to MNI space (ftp://ftp.mrc.cbu.cam.ac.uk/Colin). A two dimensional cubic spline was fitted to these points allowing interpolation between them. For each activation locus, the point on the spline (P) closest to the activation locus was determined.

Two distances were calculated to allow transformation into the unwarped space: firstly the distance between each locus and corresponding spline point P (mz) and secondly the total distance from one end of the spline to P (my). Since the vector passing through the activation locus and each point P was always normal to the gradient of the spline at P, this allowed unwarping of the anterior cingulate distribution into a rectangular space via an affine transformation. Figure 3.1 also shows the method used to unwarp the lateral prefrontal cortex loci. Successive prefrontal coronal sections of the high resolution MRI scan were examined and a series of points obtained representing the lateral cortical surface gray matter. Two dimensional splines were fitted to these points forming a set of coronal splines following the cortical surface.

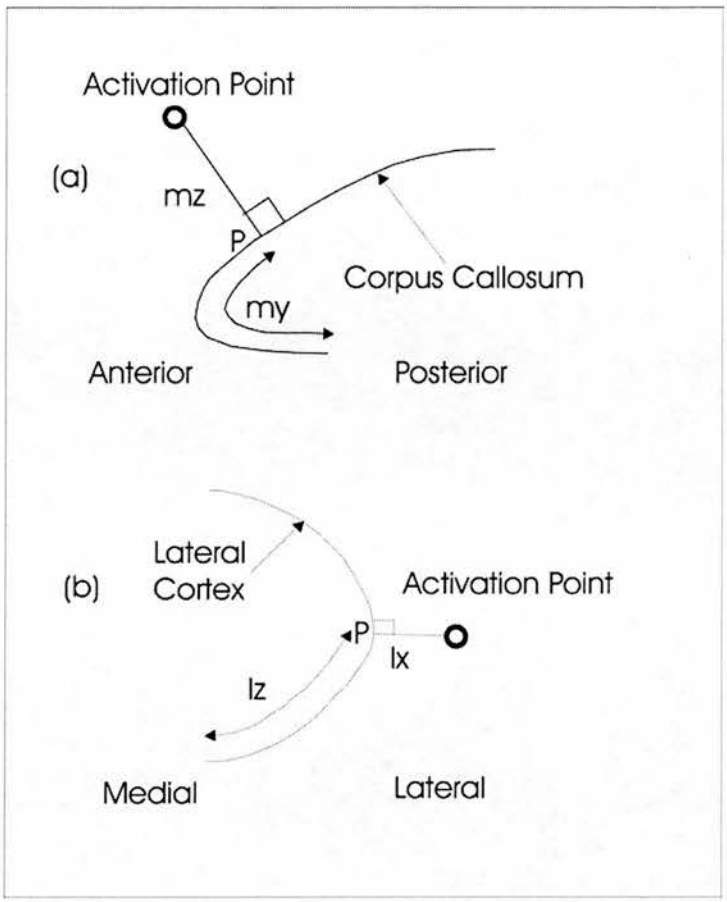


Figure 3.1 Transformations between MNI and 2-dimensional "unwarped" space medial cortex sagittal view (b) lateral cortex coronal view. Abbreviations as in text. If the warping had not been done, then the average location of an activation in the medial cortex would have been calculated as being within the corpus callosum: clearly this would not be representative of the collection of medial prefrontal cortex activation loci. Similarly, in the case of the lateral prefrontal activation loci if warping had not been done, the average activation locus would have been within subcortical structures.

A further spline surface orientated in the anterior-posterior direction was constructed from the coronal splines forming a three dimensional spline estimation of the entire lateral surface (i.e., not only at the coronal sections sampled).

For any given activation locus at a given antero-posterior location (y), a corresponding two dimensional coronal spline was identified. Next, the shortest distance between the locus and the spline was identified (lx) allowing specification of a corresponding point P on the spline. As before, the total distance along the spline to P was determined (lz). Plotting y versus lz allowed unwarping of the lateral activation loci and projection (since the shortest distance was discarded) into a rectangular space.

These calculations were implemented using custom routines written in Matlab (The Mathworks Inc.) which included the spline toolbox.

3.2.3 Hypothesis Testing and Determination of Summary Statistics

To determine whether study size causes significant bias in the observed distributions, the marginal distributions along the y, my, mz and lz dimensions were examined. The histogram of each of the unweighted distributions was compared with the corresponding distribution weighted by study size. The null hypothesis of no difference was tested with a Mann-Whitney U-test. Since these were planned *a priori* tests, a Bonferoni correction for multiple testing was not applied. If the null hypothesis was rejected, subsequent analysis was planned using weighted non-parametric methods.

It is necessary to clarify the issue of multiple testing correction. When differences between a series of pairs of variables are sought, there is an increased chance that a "significant" finding will be identified which is not genuinely so: i.e., a false positive. For example, if the threshold of significance for each test is 5% and say, six such tests are done, the probability that one test will be "significant" just by chance far exceeds 5% (Altman, 1991, p211). The most elementary correction for such multiple testing is the Bonferroni method, which involves multiplying each significance value by the number of repeated tests. Provided a small number of comparisons are done (e.g. up to five) this is reasonably accurate, but for more comparisons the correction is overly conservative (increased type 2 error).

Nevertheless, statisticians routinely advise (e.g., Altman, 1991, p211) "I do not recommend that large numbers of comparisons are performed, which would suggest poorly specified research objectives". One way to avoid a need for a multiple testing correction is to pre-specify (at the point of study planning, not once the data is being analyzed) a *small* number of statistical tests: often referred to as "*a priori* definitions". In fact, the above statement about the Bonferroni correction is unnecessary, since applying such a correction would have made it more difficult to detect a study weighting effect requiring a more complex (and therefore undesirable) weighted analysis method (described below since it was planned, and would have been used if necessary). Another method to control for multiple testing is to first apply a single "omnibus" test which determines whether a difference is likely to exist somewhere within the series of multiple comparisons. Only if the omnibus test is significant are a *few* multiple tests actually done. The multivariate tests (2-dimensional Kolmogorov-Smirnov and MANOVA) described below were planned to be used as omnibus tests in this manner, though in practice the final results were unaffected by a Bonferroni correction due to the high level of significance.

For unweighted analysis, and following a previous meta-analysis (Duncan & Owen, 2000), an initial omnibus comparison of the cognitive and emotion induction distributions was done using a 2-dimensional Kolmogorov-Smirnov test. Assuming rejection of the null hypothesis, subsequent 1-dimensional Mann-Whitney U-tests were planned to localise the difference.

For a weighted analysis, an initial omnibus comparison of emotion induction and cognitive task distributions was done using a weighted multivariate analysis of variance (MANOVA) test with the stereotactic coordinates as dependent variables, and group membership as independent variables. Canonical correlation analysis was used to calculate the MANOVA (Zinkgraf, 1983), this being based on a modified weighted version of the covariance matrix (Rousseau, 1987) and other weighted methods (e.g. weighted Pearson's correlation). Such an approach also allows (weighted) 1-dimensional t-test calculations (Knapp, 1978), which were planned if the omnibus test of the null hypothesis was rejected. Significance was determined using resampling methods (Good, 1994). The univariate tests are unaffected by

unbalanced sample sizes. The multivariate tests are similarly unaffected, particularly since only one main effect (group membership) is considered.

These calculations were implemented in Matlab using custom routines, which made use of the statistics toolbox. The U-test was obtained from a collection of available routines (http://www.biol.ttu.edu/Strauss/Matlab/matlab.htm) and a weighted canonical correlation analysis routine modified from an unweighted version in the same distribution, which had already incorporated resampling methods for significance determination. The 2-dimensional Kolmogorov-Smirnov calculation was implemented using existing C code (Press, Teukoloski, Vetterling, *et al*, 1994), which for convenience was interfaced to Matlab using a custom "mex" routine.

Assuming rejection of the null hypothesis of such tests, planned summary statistics comprised the mean and standard deviation, weighted if necessary, for each distribution in unwarped and original MNI space. In addition, calculations of 2-dimensional probability distributions of these summary statistics were planned to be overlaid for illustrative purposes on to the high resolution MRI scan. These calculations were implemented using further custom Matlab routines, which made use of existing "internal" SPM99 code (http://www.fil.ion.bpmf.ac.uk/spm/). MRIcro (http://www.psychology.nottingham.ac.uk/staff/cr1/index.html) was used to overlay the probability distributions.

3.3 RESULTS

A total of 137 relevant cognitive task studies and 44 relevant emotion induction studies was identified using the above method. Other studies did not report significant prefrontal activation loci. There was an initial 91% agreement between the authors for selection of activation loci. In the case of disagreement, virtually all were due to missing a more significant activation. In a few cases of antero-inferior activation near the sagittal plane, it was unclear initially whether the point should be assigned to the medial or lateral group of points. Inspection of the Talairach Atlas (Talairach & Tournoux, 1988) made such determination straightforward. Determination on the basis of BA was not done because of variation in opinions regarding the location of BAs in the orbitofrontal (and therefore the adjacent subgenual anterior cingulate) cortex.

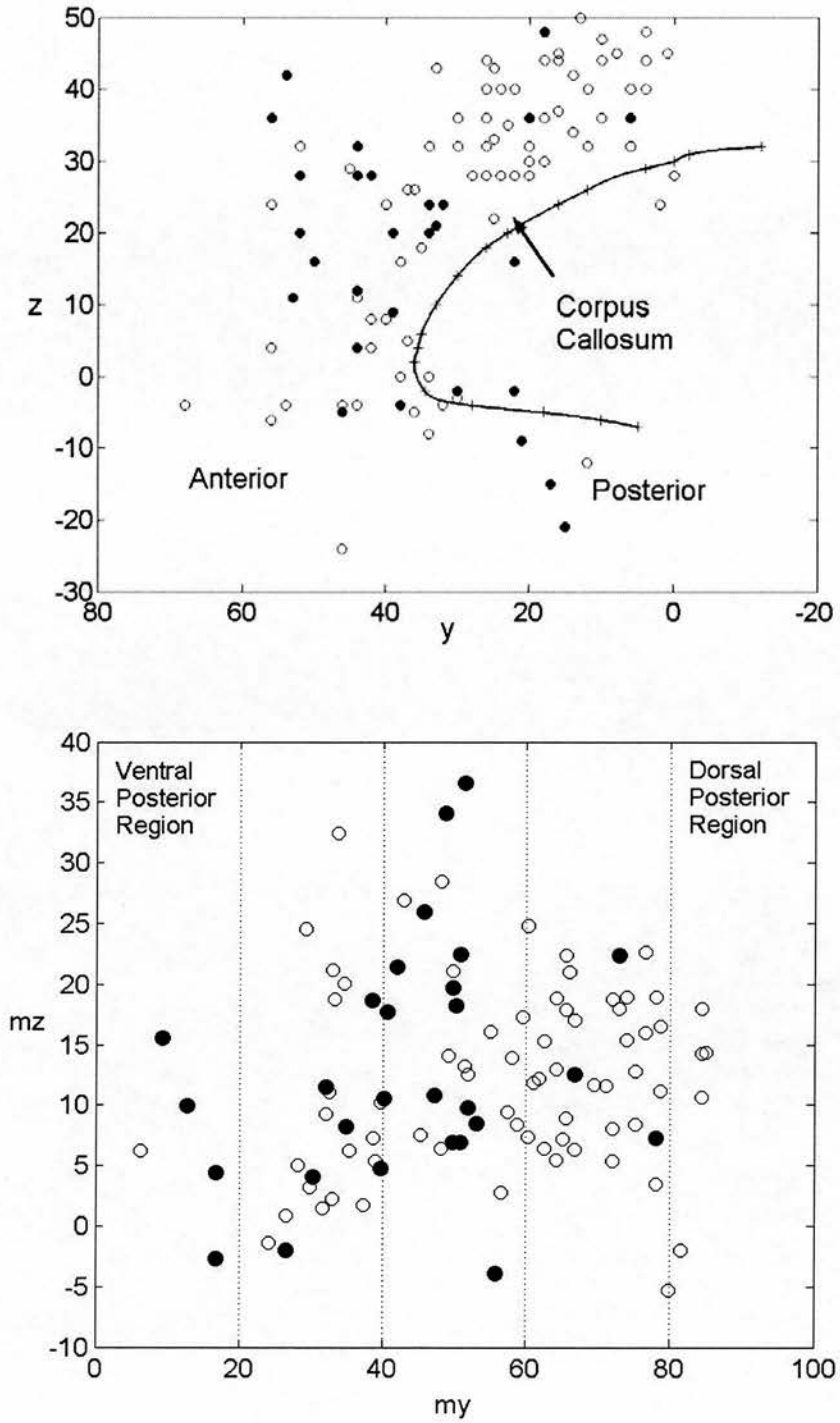


Figure 3.2 Medial cortex activation loci (open circles cognitive tasks, filled circles emotion induction)

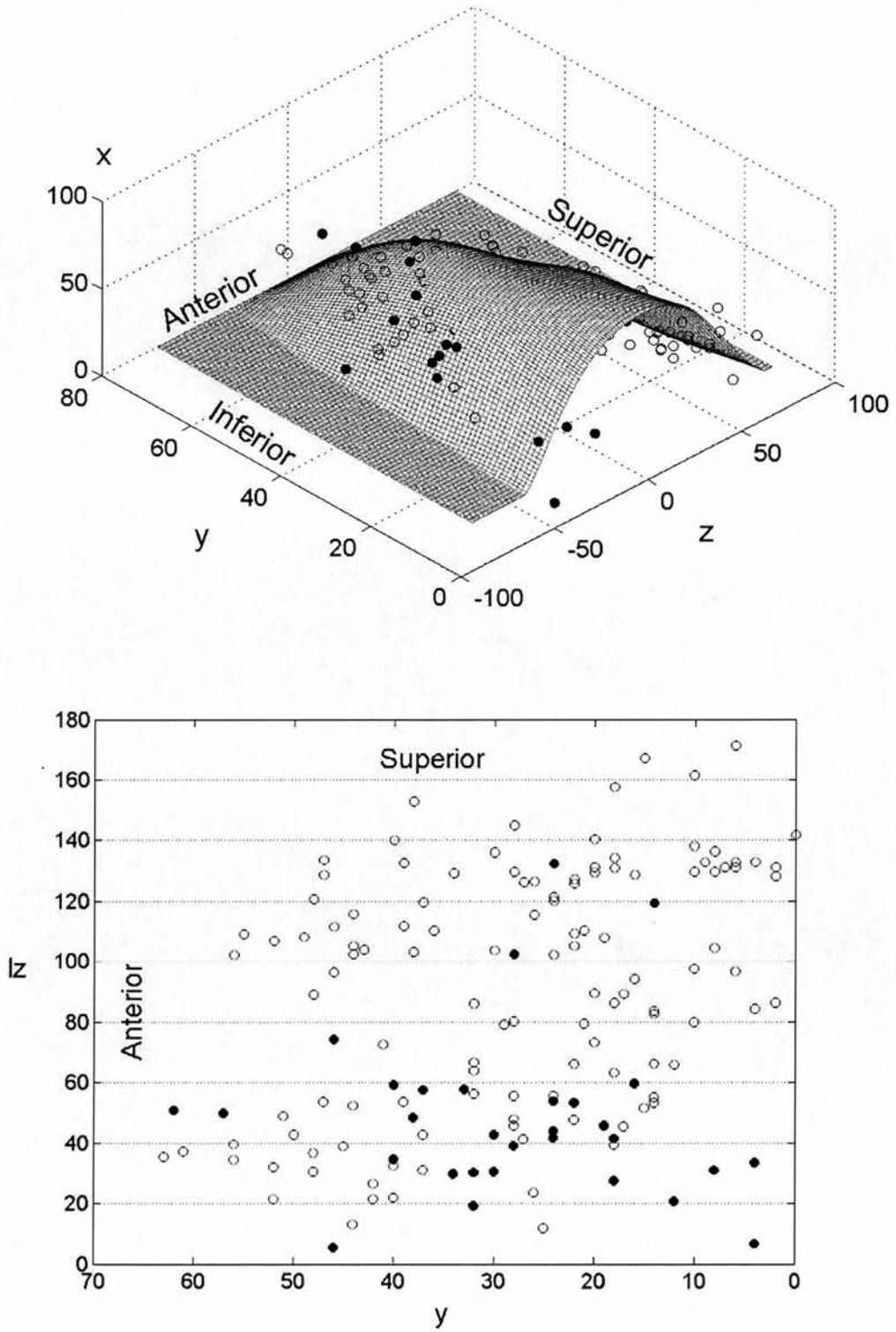


Figure 3.3 Lateral cortex activation loci (open circles cognitive tasks, filled circles emotion induction)

	Medial Cortex			Lateral Cortex		
	Transformed		MNI	Transformed		MNI
Cognitive Tasks	(52,12)	*(18,7)*	(5,28,31)	(29,91)	*(16,40)*	(54,28,18)
Emotion Induction	(42,12)	*(16,10)*	(5,46,18)	(28,48)	*(14,29)*	(42,28 ,-16)

Table 3.1 MNI and transformed mean coordinates for medial and lateral cortices. Coordinates defined as follows: medial transformed (my,mz), lateral transformed (y,lz), standard deviations in italics, MNI (±x,y,z).

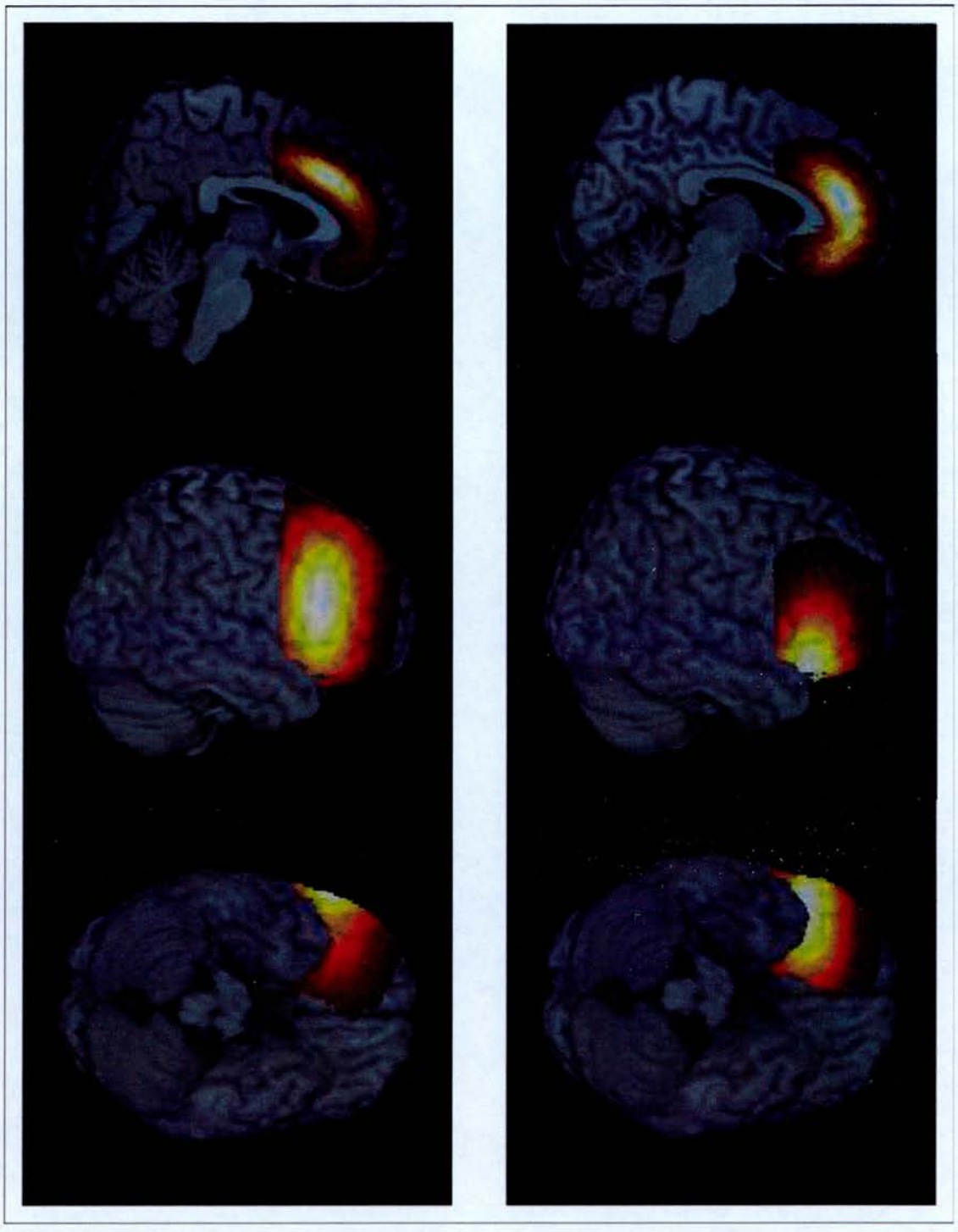


Figure 3.4 Relative probability of activation location in reviewed studies: cognitive task (left), emotion induction (right).

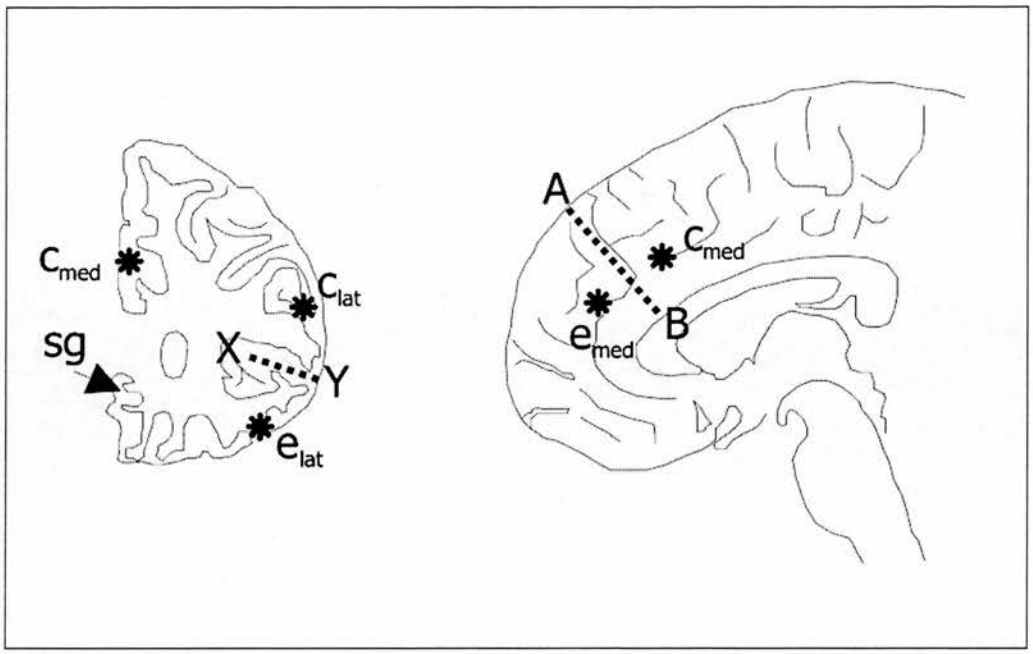


Figure 3.5 Estimated boundaries between cognitive task and emotion induction activation regions. Para-sagittal view (x=5): cognitive task (c_{med}) and emotion induction (e_{med}) mean coordinates. Coronal view (y=28): cognitive task (c_{lat}) and emotion induction (e_{lat}) mean coordinates. Plane AB, normal to sagittal plane, passing through MNI points (5,50,41) and (5,30,15). Plane XY, normal to coronal plane, passing through MNI points (42,28,2) and (56,28,-3). sg – subgenual AC

Figures 3.2 and 3.3 show the medial and lateral point distributions in original MNI and unwarped space. The reported emotion induction activation loci appear more inferior to the cognitive activation loci; however, a considerable overlap is present. The unweighted and weighted histograms of each of the unwarped marginal distributions were not significantly different ($p>0.2$ in all cases). Subsequent calculations are therefore unweighted. The difference between the 2-dimensional distributions of cognitive task and emotion induction loci was highly significant for both the medial ($p=0.00011$) and lateral distributions ($p=1.1e-9$). In the case of the medial distributions, the difference was significant in the my direction ($p=0.00010$) but not in the mz direction ($p=0.89$). In the case of the lateral distributions, the difference was significant in the lz direction ($p=1.7e-7$) but not in the y direction ($p=0.85$). Equivalent weighted calculations produced a very similar result.

Summary statistics are listed in table 3.1. Figure 3.4 shows corresponding probability distributions calculated for the original MNI space. Emotion induction tasks clearly result in a higher probability of reported inferior medial activations in comparison to cognitive tasks. However, the most posterior inferior medial cortex is not reported active as often as more rostral areas. Similarly, the lateral orbitofrontal cortex is more often reported active than medial orbitofrontal areas in emotion induction studies.

Figure 3.4 illustrates that there is a substantial overlap between cognitive and emotional probability distributions. Nevertheless, it is possible to estimate boundaries between these regions defined by planes passing through the intersection of the cognitive and emotional distributions. This is shown in figure 3.5. In the case of the medial prefrontal cortex, the boundary is supragenual. In the case of the lateral cortex, the boundary passes through the anterior insula. The coronal view also illustrates that the medial orbitofrontal cortex is continuous with the subgenual anterior cingulate (sg). As stated above, this combined region tends to be reported active less often in emotion induction tasks compared with more lateral and superior regions.

3.4 DISCUSSION

3.4.1 Summary and Conclusions

The method described here allowed extraction of activation loci from a large number of studies in a reproducible manner. The effects of variation in study size were investigated. No significant effect on the distribution of points was found. It should however be noted that the most recent published studies have not been included. Such studies have not been included because the number of such publications is increasing exponentially (Cabeza & Nyberg, 2000) and consequently, some form of sampling would be required. This was regarded as undesirable because of the large number of studies already included without sampling.

The segregation of cognitive and emotional function in the medial cortex appears in good agreement with that previously described (Bush, Luu & Posner, 2000). This suggests that possible selection biases of the types addressed here were not a significant factor in Bush's study. It is however desirable that one study does not have multiple inclusions in a meta-analysis (Stroup, Berlin, Morton, *et al*, 2000). This meta-analysis examined only the most significant (1^{st}) reported activation loci. It would be possible to include 2^{nd} and perhaps 3^{rd} most significant loci. However such points may be more susceptible to artifactual influence, including the effect of missing data (e.g., loci not reported because less significant), the effects of different clustering algorithms used for image analysis and variations in image smoothness between studies.

The identified pattern is not limited to the anterior cingulate but, as hypothesised, is present throughout the entire medial prefrontal cortex. The best estimates (average values) of most likely reported activation coordinates lie close to the cingulate sulcus. Whilst the rostral medial cortex is most likely to be reported active in emotion induction studies, unexpectedly, the most posterior inferior part of this region is infrequently active. The lateral prefrontal cortex has been similarly studied. As hypothesised, the inferior regions were most frequently reported active with emotion induction tasks, and the dorsolateral regions with cognitive tasks. Unexpectedly, the lateral orbitofrontal (BA47) region was more often reported active in comparison with the medial orbitofrontal region.

As noted earlier, studies of isocortical development (Mega & Cummings, 1997) identify two trends; a rostral emotion related region comprising the orbitofrontal, insula and temporal polar regions which extends into the rAC, and a more caudal evaluative or cognitive processing region comprising the parahippocampal and entorhinal regions, extending through the posterior cingulate to the rAC. The results of this meta-analysis, in particular the estimated boundaries between the emotional and cognitive regions (figure 3.5), appear to be in good agreement with these studies. Nevertheless, there is a substantial overlap between the emotional and cognitive processing distributions (figure 3.4).

One reason for the unexpected finding of fewer reported activations in the medial OFC and subgenual AC could be artefact. fMRI studies unlike PET, suffer from "susceptibility artefacts" which consist of localised regions of signal loss adjacent to air spaces such as sinuses (Lipschutz, Friston, Ashburner, *et al*, 2001). The frontal and sphenoid sinuses are adjacent to the subgenual anterior cingulate and posterior medial orbitofrontal cortex. However, examination of the data-set indicated that 86% of the studies used PET. Recalculation omitting the fMRI studies did not change the distributions significantly. Of course, the extracted distributions are dependent on the method used; however there appears to be no clear reason why such a distribution should occur as a consequence of artefact. If not due to artefact the distribution may reflect genuine segregation of function; e.g., the anatomical distinction between medial "limbic" and lateral OFC described by Alexander (Alexander, Crutcher & DeLong, 1990); however, that work suggested that from a functional perspective, the medial OFC was more related to emotion than the lateral OFC region, which is not consistent with this study.

The posterior subgenual AC may be homologous with the precommissural septal region in animals (Heath, 1954; Williams & Warwick, 1980) and is connected to the temporal lobe portion of the hippocampus via the fornix. The uncinate gyrus of the hippocampus includes a number of structures in the prefrontal posterior subgenual region (Duvernoy, 1991, figure 14). The medial and lateral OFC can be anatomically distinguished; the medial region having its strongest connections with the hippocampus and adjacent cingulate, the caudal lateral region having strong connections with the amygdala, insula and temporal pole (Elliott, Dolan & Frith,

2000, p308). This suggests that fewer reported emotion induction loci in the posterior medial OFC and posterior subgenual AC might be due to it being closely related to the caudal evaluative processing region described in developmental studies. However, there were not a large number of activation loci reported from cognitive task studies in this region; consequently this possibility seems less likely.

Recently, the first electrophysiological study of primate subgenual activity has been reported, which investigated whether neurons in this region respond to taste, olfactory, and visual stimuli (including faces and related movement) (Rolls, Inoue & Browning, 2003), in a similar manner to adjacent orbitofrontal neurons. Surprisingly, there was no evidence for such response. Instead, a highly significant increase in firing occurred when the animal fell asleep. If this is also the case in humans, reports of abnormal activity in depressive illness might just reflect altered alertness (Rolls, Inoue & Browning, 2003). Rolls and colleagues emphasised that there are very few brain regions known to show a substantial increase in firing rate during sleep (Rolls, Inoue & Browning, 2003). However, there is considerable evidence of increased serotonergic and noradrenergic activity in non-REM sleep (Hobson, Pace-Schott & Stickgold, 2000), and the animals in Rolls' study were only in non-REM sleep (Rolls, Inoue & Browning, 2003). Additionally, there is a well organised topography of projections from the cingulate to the brainstem in the primate (Vilensky & van Hoesen, 1981) with posterior cingulate neurones projecting to the lateral brainstem and anterior cingulate neurones projecting to the medial brainstem (Vilensky & van Hoesen, 1981). In the case of the subgenual anterior cingulate neurons, strong projections to the midline dorsal raphe nucleus have been described, resulting in speculation that the subgenual anterior cingulate may regulate the serotonergic system (Freedman, Insel & Smith, 2000). Clearly, this possibility seems even more likely, given the recent electrophysiological reports.

These findings may be relevant to the study of mood disorder, since it is accepted that sleep disturbance is very common, and manipulations such as sleep deprivation can have a significant mood elevating effect in both depressed patients and healthy controls. Further study of the interactions between the subgenual anterior cingulate and the brainstem monoamine systems is clearly indicated. In general though, the lack of imaging reports of maximal activation in the subgenual

anterior cingulate might be consistent with Rolls' electrophysiological findings. This is because there is expected to be a close match between brain regions representing rewarding and aversive experience of stimuli and brain regions active when humans are experiencing emotion (Rolls, personal communication). Further electrophysiological work is underway to investigate whether the more rostral anterior cingulate represents the rewarding or aversive aspects of stimuli (Rolls, personal communication).

Structural and functional abnormalities reported in the imaging literature for most psychiatric disorders have recently been reviewed; the dementias, learning disability syndromes, schizophrenia, mood disorders, anxiety disorders, substance misuse and eating disorders (Steele & Lawrie, 2004a, in press). As stated earlier, focal structural and functional abnormalities have been reported in the subgenual AC and OFC in both depressive illness and schizophrenia. Previously, in the context of a study on depressive illness, it has been suggested that the rAC may comprise sub-regions (Liotti, Mayberg, Brannan, *et al*, 2000). The distribution calculated here supports this distinction. It is interesting that in addition to the reported subgenual abnormalities, focal structural and functional abnormalities of the hippocampal region have also been reported in depressive illness (Shah, Ebmeier, Glabus, *et al*, 1998; Sheline, Wang, Gado, *et al*, 1996). This may reflect dysfunction of the septo-hippocampal system in depressive illness. On the basis of other evidence, dysfunction in this structure has been previously hypothesised to occur in anxiety disorders (Gray & McNaughton, 2000).

Given the limited spatial resolution of imaging studies, it is worth noting that the posterior lateral orbitofrontal region identified here merges with the anterior insula. This general area has been reported abnormally active in a number of studies on depressive illness (Drevets, 2000b). Animal lesion studies and diverse imaging studies on healthy human subjects have suggested this region is active when previously rewarded behaviour is inhibited (Elliott, Dolan & Frith, 2000). Since patients with depressive illness commonly give up previously enjoyable behaviours as a consequence of anhedonia and anxiety, there may be a link between these two observations. The dorsolateral prefrontal cortical region is often reported under active in depressive illness.

3.4.2 Future Work

The mean reported cognitive and emotional activation coordinates are of particular value for future studies of prefrontal lobe function. Not only do they assist interpretation of activation loci from patients' studies, but they may be used to define regions of interest for analyses of dysconnectivity. Chapter 5 describes an fMRI connectivity study which used these coordinates.

Having investigated the segregation of cognitive and emotional function in the prefrontal cortex of healthy subjects, a similar study of depressed subjects is now being done (forming a supervised research project for trainees in Aberdeen). Specifically, it is of interest to determine which parts of the medial and lateral prefrontal cortex are reported as maximally active, taking account of whether the depressed patient was resting in the scanner, or actively undertaking an emotion induction or cognitive task. For example, are such maximally active loci localised in the subgenual (or more rostral) anterior cingulate and ventrolateral cortex ? A different study of the anterior cingulate is underway with Chris Frith and Paul Burgess (London University), to investigate hypothesised segregation of different cognitive functions in the anterior cingulate. Similar studies of the temporal lobe are planned.

There is likely to be a limit to the extent that category specific modular functions can be localised to cortical regions: there are simply too many categories and too limited an amount of cortex. Addressing this issue for the case of object recognition in the ventral temporal cortex, Haxby and colleagues provided evidence of different categories of object being represented with distinctive patterns of response across the *entire* ventral temporal cortex (Haxby, Gobbini, Furey, *et al*, 2001; Ishai, Ungerleider, Martin, *et al*, 1999). They were then able to predict the category of object being viewed, from just the pattern of activation, with an accuracy of over 90%. This distributed organisation is in keeping with electrophysiological primate reports, and computational models of object recognition (see for references, Haxby, Gobbini, Furey, *et al*, 2001; Ishai, Ungerleider, Martin, *et al*, 1999). The method is also able to identify genuinely localised activity in response to biologically relevant objects, such as faces, which might have emerged through evolution.

Analogously, it might be possible to extend the meta-analysis method described in this chapter, to considering more than just the maximally active reported coordinate; however, there are likely to be problems with missing data, and differing image analysis methods used in the included studies.

Stochastic models of brain function, for which there is considerable evidence (chapter 4), suggests that the classical view of receptive fields (or functional segregation), in which evoked responses are invariably expressed in the same neuronal populations regardless of context, is too simplistic. Friston notes that models of brain function, which view neuronal regions as encompassing both a "bottom-up" sensory derived component, and a "top-down" predictive model (see chapter 4) component, suggests that responses evoked by sensory inputs should change with differing contexts, established by prior expectations from higher levels of processing (Friston, 2002). Consequently, when a neuronal population is predicted by top-down inputs, it will be easier to activate than when it is not (Friston, 2002). In contrast, the classical view of receptive fields assume that evoked responses will be invariably expressed in the same neuronal populations, regardless of context. Friston discusses electrophysiological and imaging examples in which real neuronal responses are not invariant, arguing that in fact such "extra-classical" effects are *common* in the brain.

From an anatomical perspective, backward and lateral afferents to a given brain region are believed to be crucial for such top-down modulatory influences. In terms of cognitive processing: "a particular region may not act as a reliable index for a particular cognitive function. Instead, the neural context in which an area is active may define cognitive function" (McIntosh, 2000). Consequently, contextually non-invariant functional segregation may be consistent with the hypothesis that the brain functions in a predictive manner. Future mapping will have to take account of these issues. Chapter 4 discusses stochastic models of brain function in more detail.

CHAPTER 4

ASSOCIATIVE LEARNING AND PHYSIOLOGICAL REGULATION

4.1 INTRODUCTION

The mechanisms of emotional regulation are increasingly recognised as relevant to the study of mood disorder (Gross, 1998). This chapter begins by introducing various basic concepts which are central to theories of *physiological* regulation: open and closed loop gain, goal states and error signals, plus characteristic features of instability in such mechanisms, and their potential relationship to clinical features of illness and disease. This leads to a discussion of Solomon's opponent process theory of affect control, which has long been argued to describe the time course of the objective and subjective emotional response to a wide variety of motivationally significant events, in humans and animals.

Imaging and transcranial magnetic stimulation (TMS) research into mood disorder have resulted in suggestions that BGTCL function is abnormal. To explore this, a simple control model of a BGTCL is described and used to interpret TMS studies of mood disorder. An adaptation of the model to allow investigation of hypothesised abnormal BGTCL connectivity in imaging studies of depressive illness is also discussed.

The remainder of the chapter focuses on various stochastic theories of brain function which are directly relevant to the fMRI study of depressive illness described in chapter 5. The discussion begins with a summary of the mathematical theories of associative learning behaviour. Importantly, these same theories also appear to describe neuronal activity in many brain regions: the evidence from animal and human imaging studies is discussed. Finally, the Kalman filter model of brain function is described in some detail, since it is the basis for the model used in chapter 5.

4.2 BASIC CONTROL SYSTEM CONCEPTS

A control system is defined as comprising a collection of interconnected components that can be made to achieve a goal (e.g., maintaining a physiological variable within set limits) in the face of external unpredictable disturbance. There are two basic configurations: "open-loop" and "closed-loop": figure 4.1. The output y is a function of both the input r, and the unpredictable disturbances δu. In the case of the *closed loop* configuration, the output y is detected by a sensor which sends a feedback signal z to the input r. An error signal e is defined

$$e = r - z \qquad 4.1$$

which provides input for the controller. In this configuration the feedback is negative.

This means that when z matches r then e is zero, but if there is mismatch, the polarity of e is such that the controller is signalled to change y in the direction of minimising the mismatch. This configuration affords stability, and is by far the most common from a physiological perspective (Guyton, 1991). The opposite is positive feedback, which tends to maximise any mismatch causing "vicious cycles". Such systems do occur in physiology, but are uncommon and usually embedded within a negative feedback system (Guyton, 1991).

Gain is defined simply as the ratio of output to input. In the case of an open loop system the open loop gain (OLG) is (Khoo, 2000)

$$OLG = Gc.Gp \qquad 4.2$$

where Gc and Gp are the controller and controlled object ("plant") gain respectively and '.' a multiplication operation. In the case of a closed loop system, the closed loop gain (CLG) is (Khoo, 2000)

$$CLG = \frac{Gc.Gp}{(1 + Gc.Gp.H)} \qquad 4.3$$

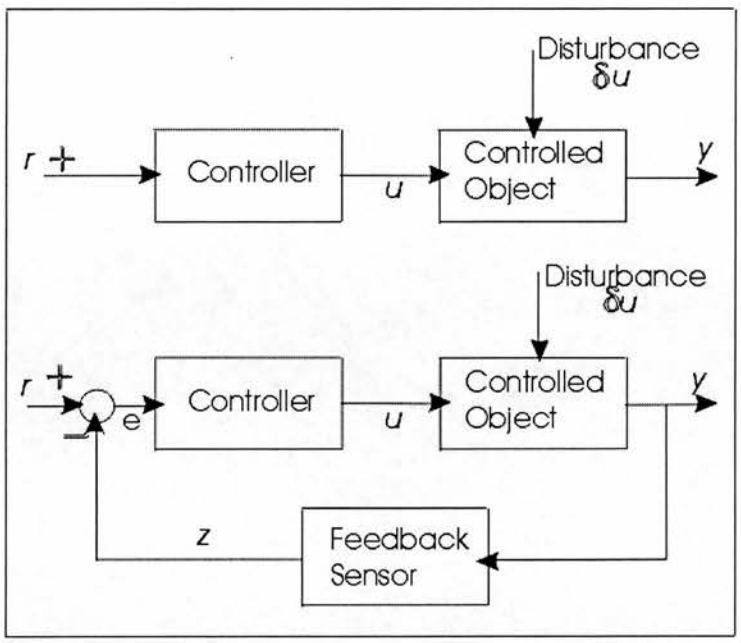


Figure 4.1 The top diagram illustrates a generic open loop control system, the lower diagram a closed loop system. δu represents unpredictable external influence which changes y. In the re-entrant closed loop configuration, it is e which is input to the controller, not r. Abbreviations: input signal (r), output signal (y), error signal (e), controller output (u), feedback signal (z).

where H is the feedback gain. This means that a closed loop system has a greater ability to maintain a regulated variable within narrower limits than an equivalent open loop system. Additionally, equation 4.3 implies that although effects of the input disturbances are reduced, they are never completely removed, resulting in a small steady-state error (unless $Gc.Gp.H$ = infinity). Characteristic instabilities occur in control systems. There are two main ways this can occur. One is high gain, the other is large signal delays. With regard to the latter, signals are not propagated instantaneously through a system, but are transferred over a finite delay period. Whilst for simplicity the above discussion defines each block as a gain comprising a simple ratio of output to input, in the more general case each block represents a "transfer function" which is realised through a mathematical operation termed a "convolution". Some of the main features of instability will be illustrated with a simple simulation. Instability is important, since various aspects of disease may reflect control instability.

To demonstrate the combined effects of feedback, gain and latency, consider figure 4.2. This comprises a simple negative feedback system with latency and gain elements, and is modified from Hung's discussion of oculomotor control instability (Hung, 2001). Calculations have been done in the Matlab Simulink environment. A "step function" is input, which is 0 for a given period then jumps instantaneously to the value 1 thereafter. This represents the neural command signal for a transition between two states, which might be two different eye directions or limb positions. The objective is for the system to move as fast and as accurately as possible from one position to the other. Figure 4.3 shows different types of system response under different conditions of gain and signal latency. In condition (a) the system moves from the first to second state, but does so over a long period. In condition (b) this movement is much more rapid and is probably close to *optimal*. In condition (c), whilst the system achieves the new state rapidly, it overshoots and oscillates about the second state before settling at the required value. In condition (d) the oscillations do not die out and the system never achieves stability. These four conditions reflect different amounts of "damping". Condition (a) is overdamped, (b) optimally damped, (c) underdamped and (d) so underdamped that the system persistently

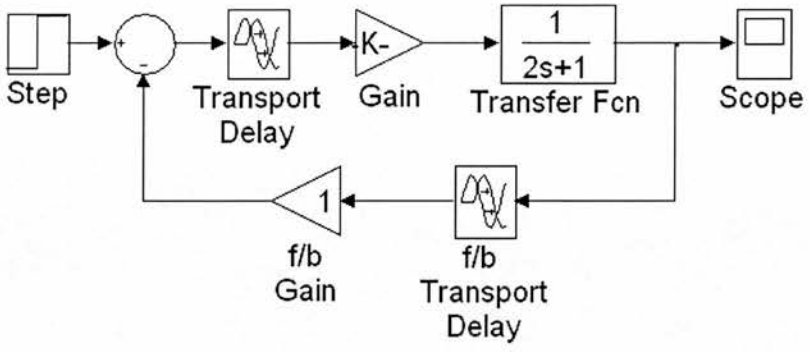


Figure 4.2 A simple negative feedback system based on detailed studies of the oculomotor system (Hung, 2001). In both the accommodation and vergence systems, the latency is long and the steady-state gain is high relative to the dynamic response of the system. A simple feedback loop would exhibit unstable oscillations (present at the "Scope" and shown in figure 4.3c and d). However, both accommodation and vergence systems separate control into two components: an initial fast open-loop component, and a slower closed loop component (Hung, 2001). The absence of initial feedback ensures stability (at the expense of increasing error) and the later closed loop control then minimises the error. In the case of the saccadic system, long latency and fast dynamics are also a feature, and it also has an initial open loop movement; however, unlike the accommodation and vergence systems, it uses later saccades for error reduction (Hung, 2001). The description, of an initial open loop movement followed by a closed loop correction, should be compared with a Kalman filter model of limb control, discussed below, which is based on a similar observation.

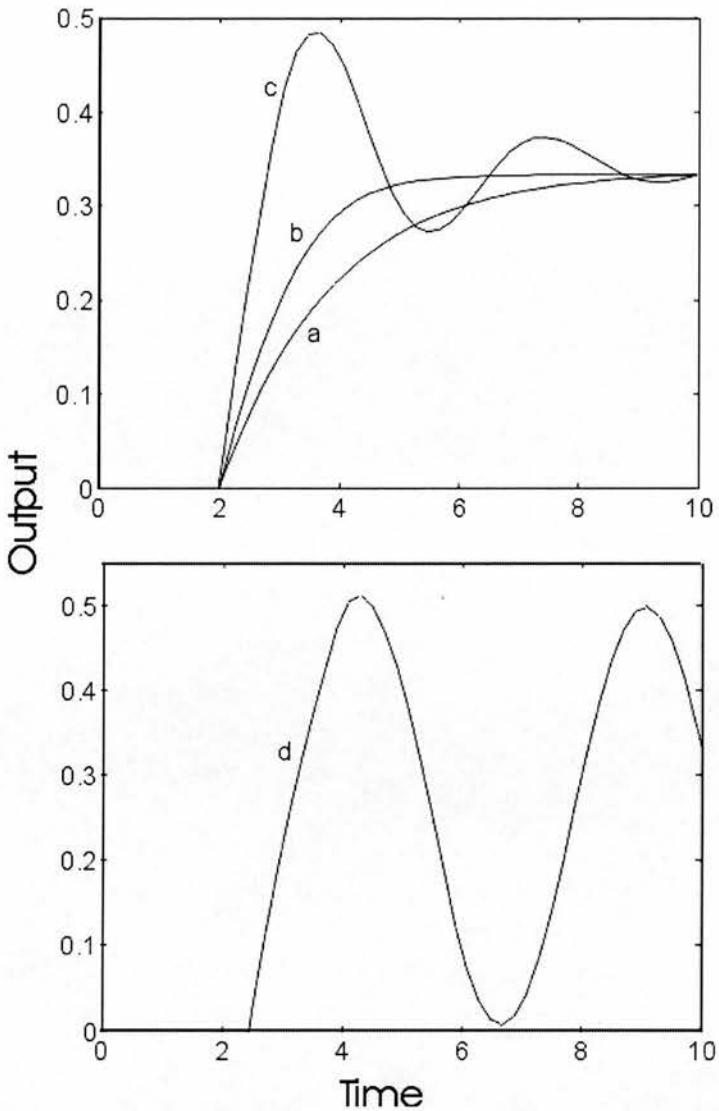


Figure 4.3 Different categories of negative feedback response under different conditions of gain and latency: (a) overdamped (gain 0.05, latency 1.0), (b) optimally damped (gain 0.5, latency 1.0), (c) underdamped (gain 2.0, latency 1.0) and (d) spontaneous and continuous oscillation (gain 2.62, latency 1.45). The units are relative to illustrate system response and not calibrated to any particular physiological system. Oscillations generally occur at high levels of gain and latency (Khoo, 2000).

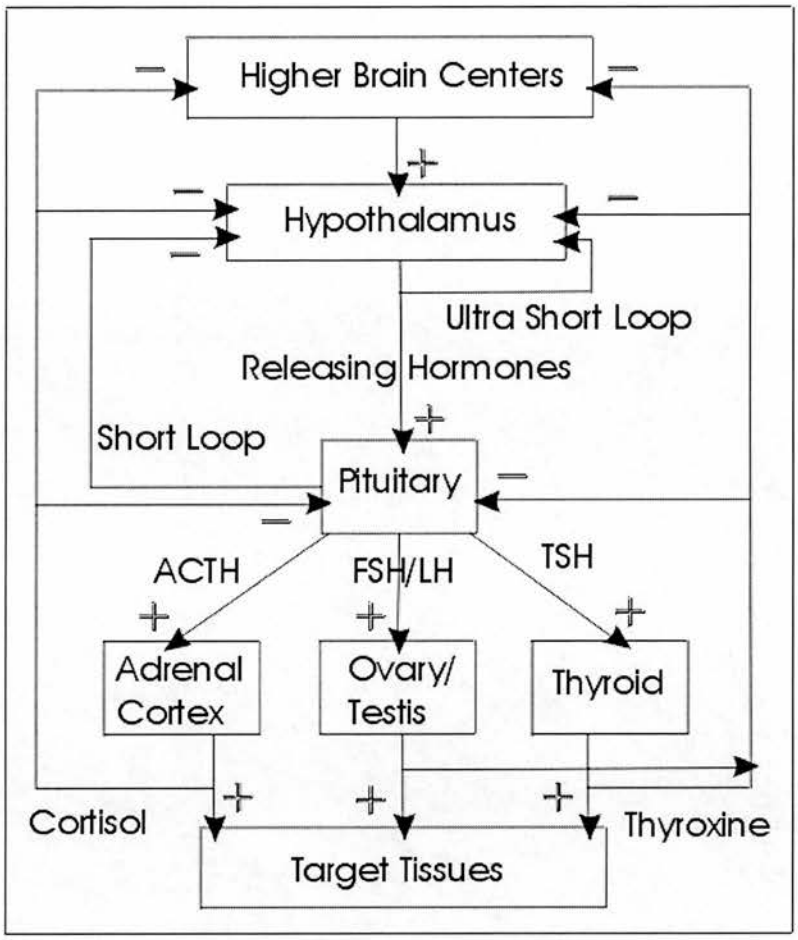


Figure 4.4 Human endocrine control system comprising many coupled hierarchical loops. Amplification (+, reflecting gain) is present, in that a very small amount of releasing hormone in the hypothalamus results in the release of a much larger amount of pituitary hormone (e.g. ACTH or TSH), which in turn leads to an even larger amount of target hormone release (e.g., cortisol or T4). This amplifying "forward" pathway is kept in check by many negative feedback controls (-) operating at all levels of the hierarchy.

oscillates. The potential relationship between these instabilities and clinical observations will be discussed later.

4.3 PHYSIOLOGICAL HOMEOSTATIC SYSTEMS

For psychiatrists, one of the most familiar examples of a control system is the endocrine negative feedback loops (eg, Hardy, 1981). These are arranged in a coupled hierarchical fashion (figure 4.4). Abnormalities of this system in depressive illness have been described (eg, Young, Haskett, Murphy-Weinberg, *et al*, 1991). Guyton discusses the homeostatic mechanisms of all the major organ systems in his popular, encyclopaedic, undergraduate textbook of medical physiology (Guyton, 1991), and defines the gain of a homeostatic system as reflecting the effectiveness with which a control system maintains constant conditions as

$$gain = \frac{correction}{remaining\_change} \qquad 4.4$$

reflecting the fact that when a change occurs, it is only partially offset by regulation (cf., equation 4.3). Control systems differ with regard to gain: e.g., the temperature control system has a much higher gain than the baroreceptor pressure system (Guyton, 1991). Another popular undergraduate text, in this case of neural science (Kandel, Schwartz & Jessell, 2000), includes a number of discussions on biological "servo-control mechanisms"; e.g., motivational states (relating to control of food or water intake, or addictive illness; chapter 51) and the functions of the hypothalamus (chapter 49). Since motivational states can be modelled as a negative feedback control mechanism, and according to Rolls there is a direct relationship between reinforcers and emotion (Rolls, 1999), this suggests that emotional response might also be modelled as a control system. Interestingly, Solomon's opponent process theory of emotion, discussed below, is such a control system theory.

Khoo's discussion of the application of modern mathematical control theory to modelling physiological mechanisms is one of the most comprehensive available (Khoo, 2000). Many models are described, which include: respiratory control (illustrating sinus arrhythmia), aortic flow, cardiac arrhythmias, neuronal dynamics

(e.g., Hodgkin-Huxley and Bonhoffer-van der Pol theories, sleep apnoea, the pupillary reflex loop, neuromuscular reflex), insulin-glucose regulation and neutrophil density regulation. Matlab Simulink programs are provided for many, facilitating understanding. The example previously discussed (figure 4.2) is a simple re-entrant loop, however realistic models of physiological function often have far more components, and involve nested feedback loops (such as figure 4.4).

Although engineering control theory can readily be applied to diverse physiological systems, there are differences; e.g., there are usually extensive cross-coupling or interaction between different physiological control systems and such systems are adaptive (Khoo, 2000). Physiological control systems usually have significant signal latencies, due for example to the time that it takes for a hormone to activate a multistep biosynthetic pathway, or the time it takes for axonal conduction (Northrop, 2000). Physiological control systems may exhibit oscillations not due to disease (as figure 4.3c or d); e.g., the ocular lens accommodation control system and fasting glucoregulatory systems (Northrop, 2000). All physiological control systems (not only neurological systems) are massively parallel, since each organ is composed of thousands or millions of cells having the same function (Northrop, 2000), which provides redundancy which is essential for robust performance in the context of injury or disease. Some physiological control systems are "push-pull", e.g., two regulatory hormones having opposite effects, such as insulin and glucagon on glucose control (Khoo, 2000), or the opposite actions of the direct and indirect neural pathways in the basal-ganglia thalamocortical re-entrant loops of the prefrontal cortex (Alexander, Crutcher & DeLong, 1990).

One of the most interesting *alleged* differences between engineering and physiological control systems concerns the generation of error signals. Both Khoo and Northrop observe that physiological control systems often do not have identifiable comparators (where an error signal is generated from the difference between the set point and the feedback signal) (Khoo, 2000; Northrop, 2000). Instead, physiological systems appear to have "implicit" comparator operations, in which a positive gain of one component is opposed to a negative gain of another, the graphical "crossing point" constituting the stable operating condition equivalent to a reference signal (Northrop, 2000). This may in part be true (opposed gain); however,

over the past decade, extensive experimental evidence has accumulated in animals, and more recently in humans, of the representation of error signals in the central nervous system (see below).

4.4 CONTROL THEORIES RELEVANT TO MOOD DISORDER

In classical behavioural theory, organisms are said to "emit" behaviour (eg, Rolls theory of emotion, chapter 2) as a consequence of stimuli applied in conditioning contexts. In control system terms, this is similar to an open loop system: stimuli are viewed as causes, behaviour as effects. However, consider a rat trained to press a lever (conditioned response) in response to a light (conditioned stimulus) through pairing the light with the availability of food (unconditioned stimulus). Powers observed that if the rat is to the left of the lever, it first moves right to achieve the lever press, and *vice-versa*. If its paw is beside the lever, it is first raised, and if it is already on the lever, the lever is just pressed (Powers, 1981, p63). Different and even *opposite* behaviours are undertaken to achieve the same end result (goal). Numerous other examples from both animal and human behaviour can be given (Powers, 1981). The explanation, according to Powers, is that such observations are consistent with behaviour operating as a negative feedback control system, with goals being represented by the input or reference signal (Powers, 1981). Powers argued that behavioural stimulus-response laws were wholly predictable within a control-system model of behavioural organisation. More subtly, it was argued that people (and animals) control perceptual input, not behavioural output. Lack of space prevents detailed discussion of Powers' model, and the reader is referred to the original texts. A different control theory, perhaps more immediately relevant than Powers' since it deals directly with emotion, is Solomon and colleagues' "affect control system". This is discussed in the next section.

4.4.1 Solomon's Opponent Process Theory of Emotion

Solomon and colleagues described a theoretical model that organises many observations associated with diverse motivational phenomena (Solomon, 1977; Solomon, 1980a; Solomon, 1980b; Solomon & Corbit, 1973; Solomon & Corbit, 1974). Figure 4.5 summarises the empirical observations forming the basis of the

system. An emotionally significant stimulus event occurs for a limited period. The vertical axis represents a measure of emotional state, the horizontal axis time. The associated cognitive-perceptual signal is represented by a square wave. A "standard pattern of affective dynamics" (Solomon & Corbit, 1974, figure 1) results, which comprises a rapid rise in intensity of emotion to an initial primary affective peak (pa), followed by an adaptive phase (d) during which a lower intensity steady state level is approached (ss). The cognitive-perceptual signal ceases with the disappearance of the stimulus event and the emotional response swings into the opposite "peak of affective after-reaction" (pb) which then slowly decays to baseline. Solomon argued that this pattern described quantifiable observed behaviour and autonomic response to a remarkably broad range of affective stimuli in both humans and animals. In humans, these stimuli include taking part in parachuting, engaging in sexual activity and taking opiates (Solomon & Corbit, 1974). In animals, these include a highly aversive experience such as electric shocks in a Pavlov harness (Solomon & Corbit, 1974). With repeated exposure to such stimuli, the affective response changes (figure 4.5). In this case the primary affective response is considerably diminished; however, the after-reaction is enhanced (Solomon & Corbit, 1974).

Solomon and colleagues sought to explain these observations with an "affect control system" (eg, Solomon & Corbit, 1973, figure 1): figure 4.6. The observed phenomena were considered to be the sum of two opponent processes, "a" and "b" of opposite hedonic valence. The "a" process is directly elicited by the cognitive-perceptual signal, and lasts only as long as this signal. In contrast, the "b" process is directly elicited by the "a" process, is of "sluggish latency", slow to build to its asymptote, and slow to finally decay (Solomon & Corbit, 1974). The effect of repeated exposure (adaptive response) is explained by the "b" process being strengthened by use and weakened by disuse, unlike the "a" process. More specifically, "b" processes approach asymptotes having values directly proportional to the quality, intensity and duration of each exposure, and inversely proportional to the inter-stimulus interval (Solomon, 1980b, p283).

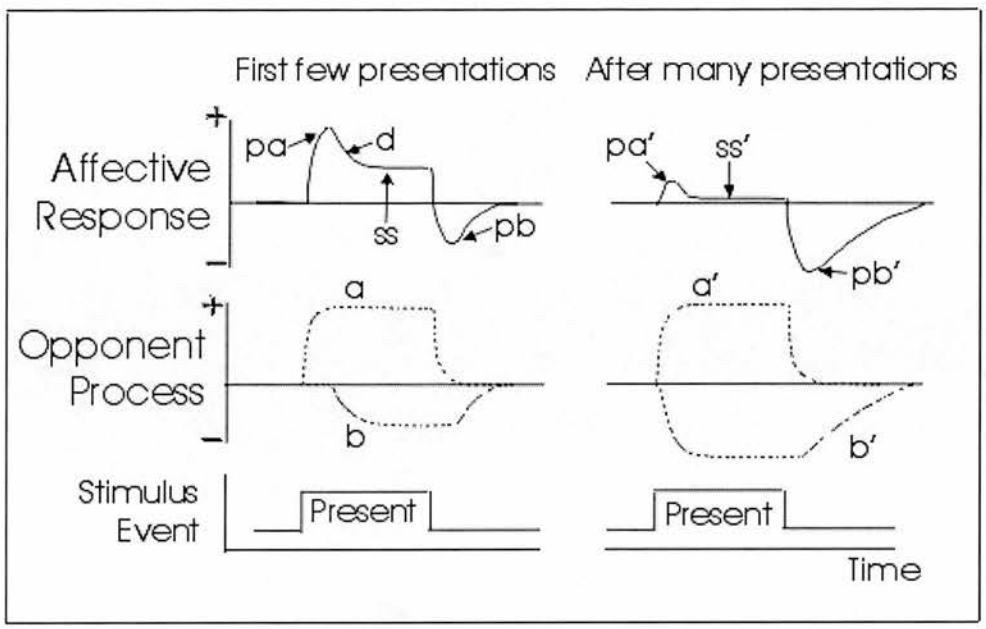


Figure 4.5 Solomon's opponent process theory of human emotion. The upper part of the diagram represents the observable (and subjective) affective response to the stimulus event shown in the lowest part of the diagram. Solomon and colleagues argued that a parsimonious theory for such an affective response comprised an unchanging "a" process, which was only present during the stimulus, and a changing "b" process, which was dependent on the "a" process (figure 4.6). The above curves are traced from the original (Solomon & Corbit, 1974, figure 1) to allow a direct qualitative comparison with that generated by a mathematical model (chapter 5). Notice that the initial affective response is represented on the left, whereas the later response after many presentations is shown on the right. In the latter case, the "b" process has become more rapid, larger (area under curve) with a slower decay to baseline. Abbreviations: peak of initial affective response (pa), decay (d), steady state (ss), peak of opposite affective response (pb), positive pleasurable emotions (+), unpleasant negative emotions (-).

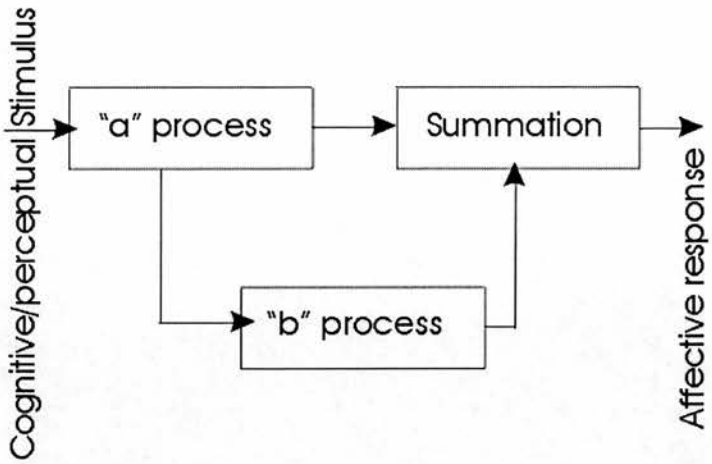


Figure 4.6 Solomon and colleagues' "affect control system". The "b" process is dependent in the "a" process, the latter only is dependent on the cognitive/perceptual stimulus. The "a" and "b" processes are of opposite valence and summed to give the observed affective response. This is equivalent to a comparator operation if the "a" and "b" processes are of the same valence (as might be the neural representation of opposite emotional states).

Notice that the control configuration suggested by Solomon includes a component for summing the "a" and "b" processes. Since he defines the "b" process as always being of opposite polarity to the "a" process, this is equivalent to a comparator if the two processes are of the same polarity. Whilst Solomon's theory is descriptive, it might be more realistic (in the case of neuronal modelling) to consider the latter situation since neurones do not exhibit negative firing rates (see though figure 4.10). Also notice that the system is not feedback, but feedforward, which is an alternative type of control (eg, Grush, 2004). Additionally, the model is non-linear and the "b" process is "non-stationary" (changes with time, in contrast to the "a" process). The *adaptive* affective response was argued by Solomon to describe features of animal bonding (imprinting in ducklings), development of long term relationships in humans, and the features of human addiction to diverse substances (Solomon & Corbit, 1974). Links with associative learning were discussed (Solomon, 1980b).

Since this work, there have been some further studies, notably in the field of substance misuse (Koob & Bloom, 1988). Recently, Daw and colleagues have developed a theory of opponent interactions between dopamine and serotonin (Daw, Kakade & Dayan, 2002), based on extensive experimental evidence of aversive events being associated with serotonergic activity, dopamine with appetitive activity, and opponent actions between the two (chapter 2). Their model, which links Solomon's theory with the temporal difference theory of dopamine action, is discussed later in the context of mood disorder (chapter 5).

4.5 DEPRESSIVE ILLNESS AND BGTCL DYSFUNCTION

4.5.1 BCTCL Control Model

BGTCLs may comprise re-entrant control systems for several reasons. Firstly, there is extensive evidence of the prefrontal BGTCLs supporting emotion, cognition and motor function (eg, Alexander, DeLong & Strick, 1986; Fuster, 1997). Solomon's theory of emotion, and Powers' behavioural theory, are clearly descriptive control models. Secondly, as described by Weiner, some symptoms of motor dysfunction, such as tremor, may reflect dysfunction of a control system

(Wiener, 1948). Thirdly, the BGTCLs are re-entrant loops, and a re-entrant structure is consistent with a closed loop control system.

Figure 4.7 shows a simple control system model of a BGTCL (cf., chapter 2). The cortex and thalamus, together with the reciprocal excitatory corticothalamic fibres, are shown as one functional unit. Input and output signal connections are indicated from this unit. The unidirectional, corticofugal, basal ganglia projections are shown as a negative feedback loop, with the pallidal output inhibiting the thalamic component of the thalamocortical unit. There is considerable evidence for segregation of cortical function being maintained throughout the BGTCLs (Alexander, Crutcher & DeLong, 1990) and so figure 4.7 is assumed to represent one such functionally segregated loop adjacent to a large number of parallel orientated loops. In accordance with a different model of the basal ganglia (Rolls, 1999, p191), each such functionally segregated loop is assumed to interact with neighbouring loops via lateral inhibition (chapter 2). In the case of the motor BGTCL, indirect and direct pathways are distinguished on the basis of differential receptor expression (e.g., D1 versus D2) (Arbuthnott, 1998), such receptors having opposite modulatory effects on loop gain.

Descriptions of other models of the basal ganglia can be found elsewhere (eg, Kropotov & Etlinger, 1999). Probably the main alternative is the "funnel" model, whereby influences from higher cognitive functions are "funnelled" towards motor areas. Rolls' basal ganglia theory is a modern version of this concept (Rolls, 1999). Such theories emphasising the anatomically convergent nature of the corticofugal basal ganglia projections have solid anatomical support; however, they are essentially open loop from a functional perspective, and can not account for the anatomical re-entrant loop structure. Additionally, they can not take account of the evidence for closed loop normal function and dysfunction in disease.

Considering the latter, as first noted by Weiner, Parkinsonian resting tremor is suggestive of high gain or latency in a control system (Wiener, 1948). However, this is apparently contradicted by the bradykinesia of movement, suggesting over-damping. A solution might be to hypothesise two interacting motor control systems; one directly controlling static postures, the other movement between postures.

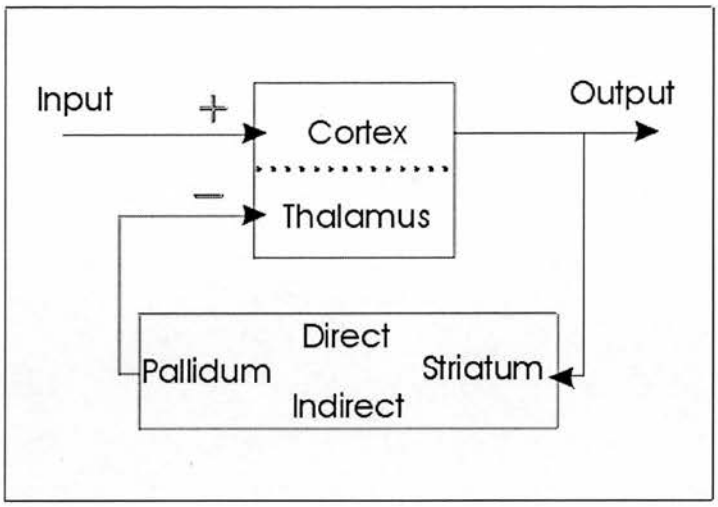


Figure 4.7 A simple negative feedback control model of a basal ganglia thalamocortical loop. This is a functional model which should be compared with Alexander and colleagues' structural model (chapter 2). The cortex and thalamus have reciprocal excitatory connections and can be considered a high gain "thalamocortical unit". The closed loop gain of the overall system is dependent on the external inhibitory basal ganglia re-entrant pathway. Depending on the gain and latencies within such a circuit, underdamped (eg, oscillatory or tremor) or overdamped (eg, bradykinetic) response could occur (see text). This model can be compared with Baev's similar hierarchical stochastic model of the basal ganglia thalamocortical loops (Baev, Greene, Marciano, *et al*, 2002). Nested feedback regulation may exist within parts of the circuit; eg, the cortex and parts of the striatum.

Such a distinction is usually made in neurology, most notably between basal-ganglia and cerebellar diseases. In this case, Parkinsonism could be viewed as a disease involving a dysfunctional high gain postural control system, and a low gain or overdamped movement control system. Parkinson's disease does not involve pathology of the cerebellum; however, it is interesting that thalamotomy, which reduces some signs of Parkinsonism, can cause the emergence of cerebellar signs (Tasker, Yamashiro, Lenz, *et al*, 1988, p306). This emphasises the interactions between different motor systems. Cerebellar disease is reflected by discoordination of movement such as dysdiadokokinesis and past-pointing. The overshoot of past-pointing suggests an underdamped control system (figure 4.3c) . There is considerable evidence for the normal action of the cerebellum being the provision of damping operations to the rest of the brain (Schmahmann, 1998). Whilst most evidence relates to motor function, cognitive and emotional functions are also implicated (Schmahmann, 1998).

The BGTCL control model predicts that Parkinson's disease (being associated with postural or static tremor) should be associated with an abnormally high BGTCL gain or signal latency. Whilst it is possible to measure neuronal signal latency in humans (thalamic response to cortical stimulation has early [100 ms, direct thalamocortical pathway] and late [300 ms, via basal ganglia] components, Kropotov & Etlinger, 1999), a literature search did not reveal reports of comparable latency studies in Parkinson's disease. Evidence for such altered closed loop gain in Parkinson's disease is included in the next section.

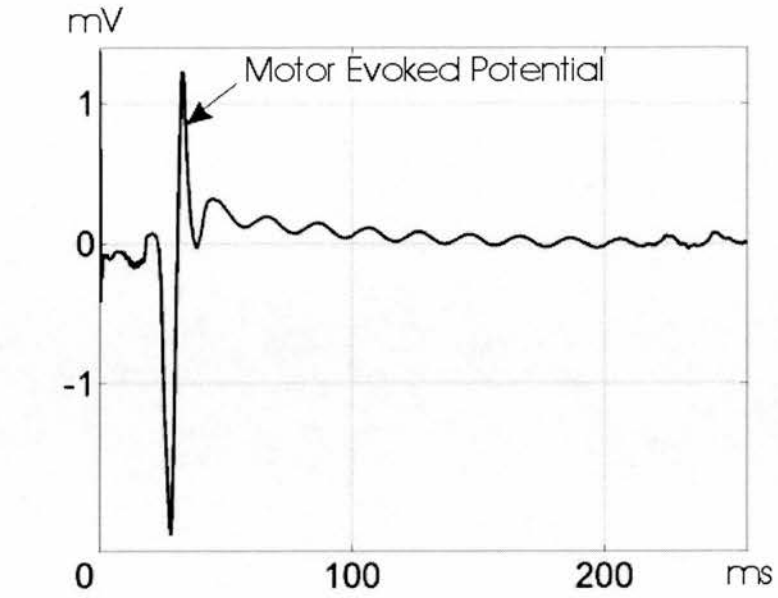
4.5.2 BGTCL Dysfunction in Depressive Illness.

Imaging and other evidence which is consistent with abnormal BGTCL function in depressive illness has been discussed (chapter 2). Transcranial magnetic stimulation (TMS) may be interpreted as providing additional evidence. This technique has been extensively applied to the study of neurological and psychiatric disorders and involves applying the field from an electromagnet to the cortex in the form of brief pulses (eg, Steele, Glabus, Shajahan, *et al*, 2000). When applied to the motor cortex of an alert individual at sufficient intensity, a single TMS pulse causes a brief volley of descending repetitive pyramidal tract discharge, activating either

directly or trans-synaptically, fast conducting fibres that project mono-synaptically to alpha motorneurones (Day, Dressler, Maertens de Noordhout, *et al*, 1989). This results in an observable movement of muscle which can be recorded as an electromyogram (EMG).

A typical EMG recording from a patient with depression is shown in figure 4.8. It comprises a motor evoked potential (MEP), which is predominately an excitatory phenomenon, followed by a prolonged period of EMG absence termed the silent period (SP), which reflects central cortical inhibitory mechanisms (Inghilleri, Berardelli, Cruccu, *et al*, 1993). Whilst a number of factors influence the SP, it is predictably related to basal ganglia dopaminergic tone: shortened in idiopathic and drug induced Parkinsonism, lengthened by L-dopa and anticholinergic drugs, and abnormally increased in Huntington's disease (Priori, Berardelli, Mercuri, *et al*, 1995).

When TMS is considered in the context of figure 4.7, the "input" comprises the TMS pulse, the "output" the EMG recording. The ratio of output to input (i.e., gain) is therefore the SP duration (or MEP amplitude) for given input TMS pulse characteristics. Thus defined, gain corresponds with the concept of "cortical excitability" (Steele, Glabus, Shajahan, *et al*, 2000): "if a given stimulus intensity is applied, the magnitude of the resulting response is dependent on the level of excitability of the cortex.". An abnormally long SP in depressive illness was found which may be consistent with two reports of reduced MEP amplitude (Samii, Wassermann, Ikoma, *et al*, 1996; Shajahan, Glabus, Gooding, *et al*, 1999) and reflect reduced (static or postural) gain in the motor BGTCL system in depressive illness. Conversely, Parkinson's Disease is associated with high BGTCL gain (high cortical excitability being indicated by increased MEPs and shortened SPs), consistent with the predictions of the BGTCL control model for postural (static) tremor.



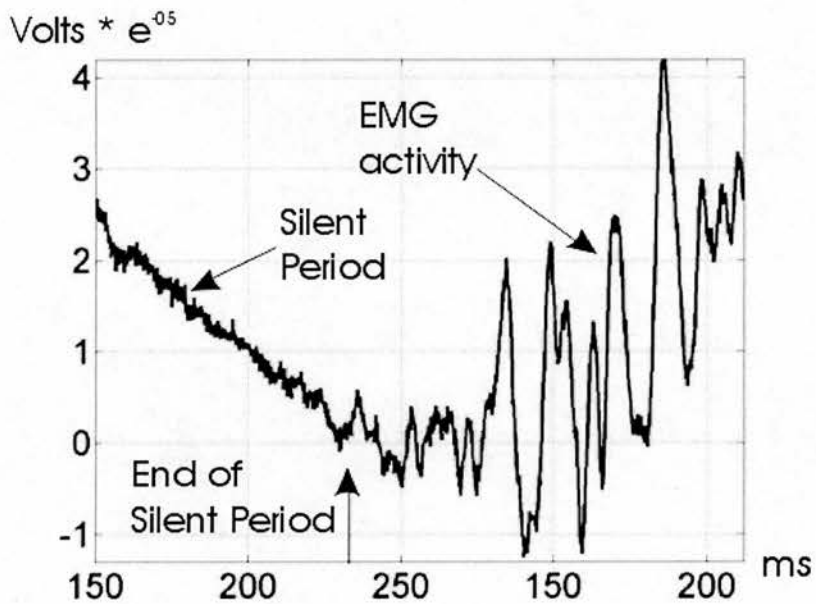


Figure 4.8 An electromyogram (EMG) recorded from abductor pollicis brevis of a depressed patient. A single TMS pulse was applied at the graph onset (0 ms) to the primary motor cortex. The top graph shows that a motor evoked potential was clearly present. The ripple is due to 50 Hz mains pickup. A bandpass filter was used to remove this artefact and the result is shown below. Baseline drift is present. Data is from a published study (Steele, Glabus, Shajahan, *et al*, 2000).

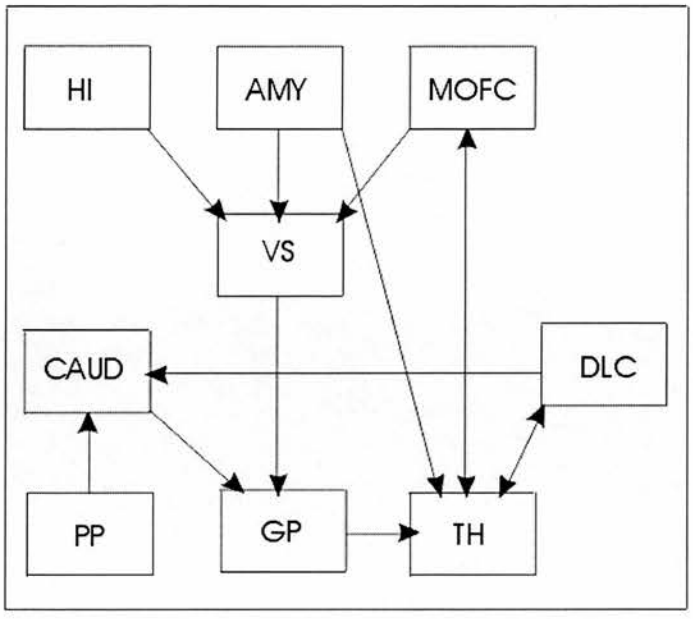


Figure 4.9 Combined limbic and dorsolateral BGTCL model used for connectivity modelling (Shajahan, Glabus, Steele, *et al*, 2002). Abbreviations: hippocampus (HI), amgydala (AMY), medial orbitofrontal cortex (MOFC), ventral striatum (VS), caudate (CAUD), dorsolateral prefrontal cortex (DLC), posterior-parietal region (PP), globus pallidus (GP), thalamus (TH). TMS stimulation was applied to a localised DLC region of depressed patients, and it was hypothesised that a change in effective connectivity would occur, not only in the DLC loop, but also in the coupled limbic MOFC loop, associated with change in mood: this was found.

Neuroimaging allows direct investigation of the components of the BGTCL system. Investigating the effects of internal pallidotomy on Parkinson's disease, Grafton and colleagues described the use of a model based on a simplified motor BGTCL structure, and reported alteration in effective connectivity (influence of one neuronal system on another) using structural equation modelling (Grafton, Sutton, Couldwell, *et al*, 1994). Based on this approach, an equivalent "limbic" BGTCL structure was proposed (figure 4.9) and used to investigate hypothesised alteration in effective connectivity in depressed patients undergoing repetitive TMS (Shajahan, Glabus, Steele, *et al*, 2002): various differences were reported. The modelling highlighted the need for an objective rationale for deciding which part of an extended cortical surface should be used to sample brain activity, due to segregation of function and partial volume effects. This led to the meta-analysis described in chapter 3. Problems with achieving a good fit between the data and model resulted in a piecewise univariate approach for determining connectivity coefficients (Shajahan, Glabus, Steele, *et al*, 2002), and the later development of the method of structural equation modelling described in chapter 5.

4.6 QUANTITATIVE ASSOCIATIVE LEARNING

4.6.1 Learning Theories

There are various limitations of the above theories. Firstly, feedback often takes too long. Secondly, the environment is only partly observable and controllable (Baev, 1998, p52). The degree of observability is usually high for relatively simple controlled objects; complex objects usually have low observability. The latter occurs when there are many environmental influences on the object for which the control system can not account (Baev, 1998). Adding additional sensory measurements may offset the problem at the expense of increased computation. The most significant difficulty associated with limited observability is that the received information is noisy (partially obscured by unpredictable external influences). Incomplete controllability refers to situations in which a single action is insufficient to move the environment from the current state to the required final state (Baev, 1998), and a sequence of actions are required. Thirdly, organisms adapt to changes in the environment, recover from damage and learn, and these features are not properly

addressed by deterministic theories. Quantitative associative learning theories do address these issues, and whilst not always clearly expressed as control theories, can often be formulated as such. Just like deterministic control theories, an error signal is a central feature of associative learning theories; however, in the latter case the error signal reflects prediction error.

Associative learning allows animals and humans to predict the occurrence of important events. Learning occurs when the actual outcome differs from the predicted outcome, the difference between the two constituting a predictive error (Schultz & Dickinson, 2000). Formally, such predictive error based learning is manifest in classical and instrumental conditioning. However, it is a central process of adaptive and intelligent behaviour in both animals and humans in general: "the notion of a prediction error relates very intuitively to the very essence of learning" (Schultz & Dickinson, 2000, p476). Learning can be viewed as the acquisition of predictions of outcomes (reward, punishment, behavioural reactions, external stimuli, internal states). Outcomes of a magnitude or frequency different from that predicted modify behaviour in a direction that reduces the discrepancy between outcome and prediction. When the outcome is perfectly predicted, prediction error is zero, and no learning occurs. The Kamin blocking phenomenon demonstrates that simple pairing of a stimulus and reinforcer is not sufficient for learning (Kamin, 1969): learning is restricted to unexpected stimuli or outcomes, and precludes learning about redundant stimuli preceding outcomes already predicted by other stimuli (Schultz & Dickinson, 2000, p476). Prediction error features in associative learning theory in two distinct ways: direct learning (Rescorla & Wagner, 1972) and indirect learning through attentional allocation (Pearce & Hall, 1980).

The *Rescorla-Wagner* learning rule proposes that the increment in the associative strength of a signal ΔV during a learning episode is directly determined by the prediction error

$$\Delta V = \alpha.\beta.(\lambda - \Sigma V) \qquad 4.5$$

where α determines the rate of learning, β the properties of the stimulus and reinforcer, λ the strength of associations with the reinforcer required to fully predict

the reinforcer and ΣV the combined associative strength of all signals during a learning episode (Schultz & Dickinson, 2000). During learning, incomplete prediction of reinforcement occurs; consequently the error term $(\lambda - \Sigma V)$ is positive and associative strength increases. Conversely, if reinforcement is omitted the error term is negative and associative strength decreases, corresponding to forgetting or extinction (Schultz & Dickinson, 2000). In their original publication, Rescorla and Wagner cite numerous observations of animal behaviour as being consistent with their theory (Rescorla & Wagner, 1972).

In contrast, the *Pearce-Hall* attentional learning rule equates changes in the associability parameter ε of the signal with the *absolute* value of the prediction error

$$\Delta\varepsilon = |\lambda - \Sigma V| \qquad 4.6$$

Similar to the Rescorla-Wagner rule, such changes are summated during such learning. However, the prediction error plays no direct role in changes in associative strength which are a function of change in associability (Schultz & Dickinson, 2000). Behavioural studies indicate that there is evidence for both types of learning (Schultz & Dickinson, 2000).

Such theories are compatible with understanding of biological neuronal function (eg, Hebb, 1949) in which predictive learning was hypothesised to result from strengthening of synaptic influence via conjoint pre and post-synaptic activity. Such neurobiological theory formed the basis of later artificial neural network models, such as those utilising the *Delta learning rule* (Schultz & Dickinson, 2000, p478). This rule is based on the least mean square error control system, developed by *Kalman* and others (Schultz & Dickinson, 2000, p478). Beginning with an analysis of Kandel and colleagues' Nobel Prize work on Aplysia (sea slug) conditioning (Hawkins & Kandel, 1984), it has been proposed (McLaren, 1989) that the prediction error could be computed by a negative feedback control system which conforms to a direct calculation of prediction error in the form of the Rescorla-Wagner behavioural rule (Schultz & Dickinson, 2000). The error term is formally equivalent to the prediction error described by associative learning rules (Schultz &

Dickinson, 2000, p479). A more complicated neural assembly would be required to implement an attentional learning rule (Schultz & Dickinson, 2000, p479).

The *temporal difference* (TD) theory was also originally inspired by behavioural data on how animals learn predictions (Sutton & Barto, 1998). Subsequently, it has been used in many engineering and computing applications that seek to solve prediction problems analogous to those faced by animals and humans (Schultz, Dayan & Montague, 1997, p1595). It can be shown (Seymour, O'Doherty, Dayan, *et al*, 2004a) that the TD error is

$$\delta(t) = r(t) + n(t) - n(t\text{-}1) \qquad 4.7$$

where $n(t)$ is the predictive value of a reward and $r(t)$ is the reward, both at time t. Predicting the sum of future rewards is a generalisation over static learning rules such as the Rescorla-Wagner, although the method by which the brain achieves this is not particularly well understood (Schultz, Dayan & Montague, 1997).

The final model considered here is the *Kalman filter* (Dayan, Kakade & Montague, 2000). This theory formalises the predictive relationship between conditioned stimuli and reward, together with how this predictive relationship is expected to change over time (Dayan, Kakade & Montague, 2000). Importantly, in addition to the prediction made for each stimulus, there is an estimation of uncertainty in the prediction, which is large if the stimulus has not been present in the recent past, and low if there is experience of a substantial relationship between stimulus and outcome (Dayan, Kakade & Montague, 2000). If the delivery of conditioned stimuli at time t is represented by $\mu(t)$, the delivery of a reward represented by $r(t)$, θ is an estimate of a weighting of association, then $\eta(t)$ reports the error in the current predictions

$$\eta(t) = r(t) - \mu(t).\theta(t) \qquad 4.8$$

and "is exactly the same error term that appears in many supervised learning rules, from Rescorla-Wagner rule for conditioning, to the backpropagation rule for [artificial] neural networks and in a slightly modified form the temporal difference

error" (Dayan, Kakade & Montague, 2000, p1220). The Kalman filter model is discussed in more detail below, and was used in the study of depressive illness described in chapter 5.

4.6.3 Neural Predictive Error Signals

4.6.3.1 Animal Studies

Schultz and Dickinson have reviewed the extensive experimental evidence for the brain exhibiting predictive error signals (Schultz & Dickinson, 2000). These include the midbrain dopaminergic, noradrenergic and acetylcholinergic systems, dorsolateral and orbitofrontal cortex, striatum, visual cortex, superior colliculus and cerebellum.

Figure 4.10 shows the activity of midbrain dopaminergic neurones during conditioning (Schultz & Dickinson, 2000). Before learning (figure 4.10, A), when an unexpected reward is delivered there is a transient increase in activity. If it is learned that an event (conditioned stimulus) which precedes the reward predicts its occurrence (figure 4.10, B), the transient increase in dopaminergic firing previously occurring just after reward delivery no longer occurs, but instead "moves backward in time" to just after the conditioned stimulus. Once such learning has occurred, if the reward is not delivered, a transient decrease in dopaminergic firing occurs just after the predicted time of the omitted reward (figure 4.10, C). This pattern of neuronal firing conforms to an error in the prediction of reward which is the error term in associative learning models (Schultz & Dickinson, 2000). The fact that omitted rewards induce opposite changes in dopaminergic firing compared with unpredicted rewards, suggests an error term that directly controls learning, rather than an attentional model (Schultz & Dickinson, 2000).

Noradrenergic neurones respond to unpredicted but not predicted rewards (Schultz & Dickinson, 2000). A similar response to aversive stimuli has not yet been explored. Acetylcholinergic nucleus basalis of Meynert neurones respond to unpredicted omitted rewards, or following performance error. The neuronal populations involved may be different.

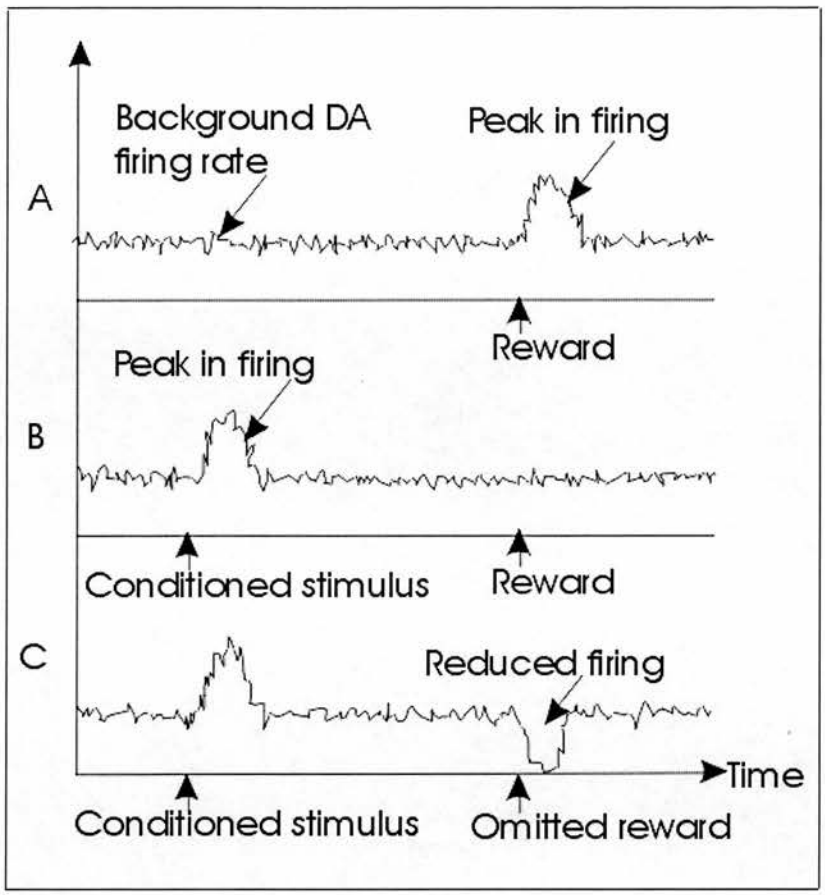


Figure 4.10 Schematic diagram showing the change in midbrain dopamine (DA) firing during associative learning (Schultz, Dayan & Montague, 1997; Schultz & Dickinson, 2000; Schultz, Tremblay & Hollerman, 1998). In A, which shows the situation before learning, a reward is unexpectedly presented and there is a transient increase in dopaminergic firing. B shows the firing pattern after learning that a stimulus regularly occurs before the reward (i.e., predicts the future occurrence of the reward): the increase in firing no longer occurs before reward delivery, but "moves backward in time" to just after the now conditioned stimulus. C shows that when the predicted reward is not delivered, a transient reduction in dopaminergic firing occurs at the time when reward should have been delivered. As noted by Schultz and colleagues, this pattern conforms to the error term in quantitative associative learning models. Other monoamines, such as noradrenaline, exhibit similar, but not identical firing patterns. These are "broadcast" to widespread brain regions as a global error or reinforcement signal, affecting synaptic plasticity. Many other non-monoaminergic brain structures exhibit similar activity (see text).

The orbitofrontal cortex neurones code reward unpredictability but further studies are required to explore the range of response. In contrast to dopaminergic neurones, many tonically active striatal neurones respond with decreased activity to primary rewards and reward predicting stimuli (Schultz & Dickinson, 2000). The first reported error driven teaching signal in the brain involved the Purkinje neurones in the cerebellum (Marr, 1969). Extensive evidence for these signals being involved in conditioning to aversive signals has been obtained (Schultz & Dickinson, 2000), and such signals may conform to the delta rule (Schultz & Dickinson, 2000). Anterior cingulate, dorsolateral prefrontal, orbitofrontal, superior colliculus and frontal eye field neurones exhibit activity which appears to be consistent with the Delta rule or Rescorla-Wagner equation (Schultz & Dickinson, 2000, p478).

The experimental evidence from these and many other studies suggests that error signal representation differs in different brain regions and systems. Many brain regions do not respond exactly in the same manner as dopaminergic neurones. Some regions respond more to aversive stimuli; e.g., the cerebellum (see also chapter 2 on the anterior cingulate). Additionally, there is evidence for hierarchical activity e.g., dopamine neurones emit a reward teaching signal without indicating the specific reward, striatal neurones adapt expectation activity to new reward situations, and orbitofrontal neurones process the specific nature of rewards (Schultz, Tremblay & Hollerman, 1998, p421). It should be emphasised that the experimental data is incomplete, and more studies are required to fully characterise the types of predictive error signals in different brain regions. Nevertheless, there is solid experimental evidence for predictive error signals in many brain regions. The following section discusses the recent replicated imaging evidence, for such signals being present in healthy humans.

4.6.3.2 Human Imaging Studies

Berns and colleagues noted that certain classes of stimuli (eg, food and drugs) are highly effective in activating reward regions, and reported that the predictability of the reward delivery modulated human nucleus accumbens (chapter 2) and medial orbitofrontal cortex activity (Berns, McClure, Pagnoni, *et al*, 2001). Pagnoni and colleagues reported a significant change in ventral striatal activity when an expected

fruit juice stimulus was withheld, concluding that human ventral striatal activity was time-locked to errors in reward prediction (Pagnoni, Zink, Montague, *et al*, 2002), as in animals (figure 4.10, C). Ploghaus and colleagues studied the effects of unexpected aversive events, reporting that unexpected painful heat was associated with activation of the human hippocampus, superior frontal gyrus, cerebellum, and superior parietal gyrus, and discussed these results in terms of the Rescorla-Wagner, temporal difference and Pearce-Hall theories of associative learning (Ploghaus, Tracey, Clare, *et al*, 2000). O'Doherty and colleagues investigated reward related learning, reporting that the temporal difference prediction error signal, which shifts backward in time during associative learning in animals (figure 4.10, A to B), also displays the same shift in the orbitofrontal cortex and ventral striatum of humans (O'Doherty, Dayan, Friston, *et al*, 2003). McClure and colleagues differentiated between unpredictability of the time and the amount of reward using a passive conditioning task, reporting positive and negative prediction errors in reward delivery time correlated with altered activity in the human striatum (McClure, Berns & Montague, 2003), in accordance with reports from studies on animals. A further study reported a temporal difference error signal in human medial prefrontal and orbitofrontal cortex, nucleus accumbens, caudate, putamen and hippocampus, during a game paradigm (McClure, Li, Cohen, *et al*, 2004), which is a more complex activity than simple conditioning. Seymour and colleagues reported a strong activation of the human bilateral ventral putamen using temporal difference error signal modelling during an aversive painful task (Seymour, O'Doherty, Dayan, *et al*, 2004b). They noted that since activity in the same structure also correlates with reward, appetitive and aversive information appear to be combined leading to motivationally appropriate behaviour in the light of both. Glascher and Buchel reported that the Rescorla-Wagner model predicted activation in the human amygdala during an implicit classical conditioning task involving acquisition and extinction on a trial by trial basis with facial stimuli (Glascher & Buchel, 2004). Finally, Lavric and Wills used low resolution electrical tomography to investigate the effects of an emotion (surprise) on human learning in relation to various formal learning theories (Lavric & Wills, 2004).

In summary, it seems reasonable to conclude that there is considerable experimental evidence for predictive error signals in many regions of the human brain, these signals being described by various formal mathematical theories.

4.6.4 Kalman Filter Models of Brain Function

4.6.4.1 General Models

Various brain regions and functions have been suggested as conforming to a Kalman filter. For example, based on experimental evidence for predictive error signals, the combined action of the amygdala, hippocampus, dopaminergic system and ventral striatum has been modelled as a Kalman filter (Dayan, Kakade & Montague, 2000). Based on anatomical structure and experimental findings, the hippocampus alone has been suggested to function as a Kalman filter (Bosquet, Balakrishnon & Honavar, 1999). Gray's theory of septo-hippocampal function is not expressed as a Kalman filter model; however, it does comprise a re-entrant system involving a comparison of predicted sensory input (or goals) with actual input (chapter 2). Wolpert and colleagues reported that a Kalman filter was able to parsimoniously model sensorimotor experimental data, and suggested that this may reflect actual neuronal function (Wolpert, Ghahramani & Jordan, 1995). Baddely and colleagues investigated a visuomotor task in humans, undertaking a "system identification" procedure to determine which of a series of quantitative models best described behaviour, and compared this with a Kalman filter model of *optimal* performance (Baddeley, Ingram & Miall, 2003): they reported that a modified Delta rule provided the best fit. Glasauer and colleagues have described a Kalman filter-like closed loop control system of vestibular function, which was argued to be consistent with experimental findings from studies on complex inertial stimuli (Glasauer & Merfield, 1997). A Kalman filter based model of human spatial orientation has been described which appears to be consistent with experimental data (Borah, Young & Curry, 1988). Additionally, a Kalman filter model has been used to demonstrate how static and dynamic events in the visual environment can be learned and recognised, given only the input images, and how a certain form of attention may be an emergent property of the interaction between top-down expectations and bottom-up signals (Rao, 1999): consistency between the theory and

animal and human experimental data has been claimed; also see another discussion (Eliasmith & Anderson, 2002). Grush has described negative feedback Kalman filter models of diverse cognitive and motor function (Grush, 2004): see next section. In summary, a number of brain regions have been modelled as Kalman filter processes and consistency with experimental findings reported.

4.6.4.2 Grush's Kalman Filter Theory and Hierarchial Models

Although most of the earlier discussion argued for various brain functions (emotion, cognition, sensory and motor function) operating in a closed loop negative feedback manner, there is evidence for open loop ("forward") function. Such evidence may not be contradictory. For example, Wolpert and colleagues described a model which could estimate a person's body state during movements (Wolpert, Ghahramani & Jordan, 1995). The experimental data comprised subjects' estimates of the position of their hands after varying periods when deprived of visual feedback. They reported that as the duration of movement increased from 0.5 to 1 sec the overestimate increased, but thereafter it decreased, independent of assistive, resistive or no external force (Wolpert, Ghahramani & Jordan, 1995). They argued that neither pure sensory inflow nor motor outflow based models could account for this, but a Kalman filter model could. Due to finite nerve conduction times, propioceptive and kinaesthetic feedback is not initially available, and the "state estimate" (hand position) is based almost entirely on an open loop "forward" model. Initial error increases with increasing time in the absence of correction signals; however, as peripheral feedback becomes available after about 1 sec, this is used to make corrections, and estimation error consequently decreases (Wolpert, Ghahramani & Jordan, 1995) (cf, oculomotor function, figure 4.3).

Building on such work, Grush has presented a Kalman filter based "emulation theory of representation" which appears to synthesise a wide variety of representational functions of the brain focusing on motor control, imagery and perception (Grush, 2004). Central to the theory is the hypothesis, that in order for the brain to engage with the wider body and environment, it constructs neural circuits that act as models of body and environment (Grush, 2004). During sensorimotor engagement, these models are driven by "efference copies" (of motor commands to

muscles) which provide expectations of the sensory feedback, and provide "top-down" processing of sensory information. The models can also be run "off line" in order to produce various cognitive phenomena: imagery, the estimation of possible outcomes, and the development of motor plans (Grush, 2004).

In addition to focusing on the above cognitive/motor function, Grush also discusses Damasio's theory of emotion, arguing that Damasio actually posits a "visceral emulator", whose function is to provide mock emotional/visceral input; i.e., emotional imagery, which might provide the substrate of emotional expectations (Grush, 2004). Other theories which Grush believes are compatible with his Kalman filter theory of emulation include: theory of mind phenomena (Frith, 2003) and cognitive linguistics (see for discussion Grush, 2004).

The computational model of brain function that Grush proposes is shown in figure 4.11. This is a general stochastic control model which blends pseudo-closed loop control and a Kalman filter. Notice that the signal sent to the emulator is a copy of that sent to the controlled object (e.g. musculoskeletal system) and is therefore an efference copy. A Kalman filter was used in the study described in chapter 5 on depressed patients: consequently, it is necessary to provide more information on the operation of the filter. Grush provides a detailed description (Grush, 2004) of the mathematical processes (Kalman, 1960) represented in figure 4.11. The particular version of the Kalman filter considered here is the discrete linear form: other forms exist, such as generalisations to continuous and non-linear systems.

The Kalman filter provides estimates of system variables (e.g. changing aspects of the environment) by using statistical models to weight each new measurement relative to past information. It also determines up-to-date uncertainties of the estimates. Providing the assumptions underlying the particular type of filter are met (e.g., linear gaussian processes), the filter is theoretically *optimal*; i.e., it is not possible to achieve a better performance. Consequently, if it is argued that a particular brain region (e.g. hippocampus), or brain function (e.g. motor function) can be modelled as a Kalman filter, this may be equivalent to arguing that the particular brain region or function adapts towards a theoretically optimal function. Notice also that if the system remains at a constant value, a Kalman filter reduces to a sequential form of classical least squares (with a weight matrix equal to the inverse

of the measurement noise covariance matrix, http://www.cs.unc.edu/~welch/kalman/Levy1997).

The Kalman filter is a multiple-input, multiple-output digital filter which can optimally estimate in real-time the states of a system based on noisy (unpredictable external influence) outputs. These states are all the variables needed to completely describe the system behaviour as a function of time (e.g., observables in a conditioning experiment) and it is possible to regard such outputs as a multidimensional signal plus noise, with the system states being the desired variables. The Kalman filter, then, *filters* the noisy measurements to estimate the desired, partially hidden, signals. Such estimates are optimal in the sense that they minimise the mean squared estimation error.

The system state is typically a series of scalar variables (vector) rather than a single variable (though the latter was the case in chapter 5), and so the state uncertainty estimate is a covariance matrix, with the diagonal terms comprising the variance of the given variable, and the off diagonal terms the covariances between any pairs of variables. Such multiple measurements are also vectors in a recursive (self-repeating) algorithm comprising two steps: the "time update" and the "measurement update" (Grush, 2004). Starting with an initial predicted state estimate, and an estimate of associated covariance, the filter calculates the weights to be used when combining the estimate with the first measurement vector to obtain an updated best estimate.

Next, the filter projects the updated state estimate and associated covariance to the next measurement time. The actual state vector is assumed to change with time according to a linear transformation plus independent random noise. As the measurement vectors are recursively processed, the state estimate's uncertainty will usually decrease because of accumulated information from past measurements. However, uncertainty increases in the prediction step, so the two influences tend to balance with time at a finite value. A Kalman filter is therefore a digital filter with time varying gains, the latter being collectively referred to as the "Kalman gain".

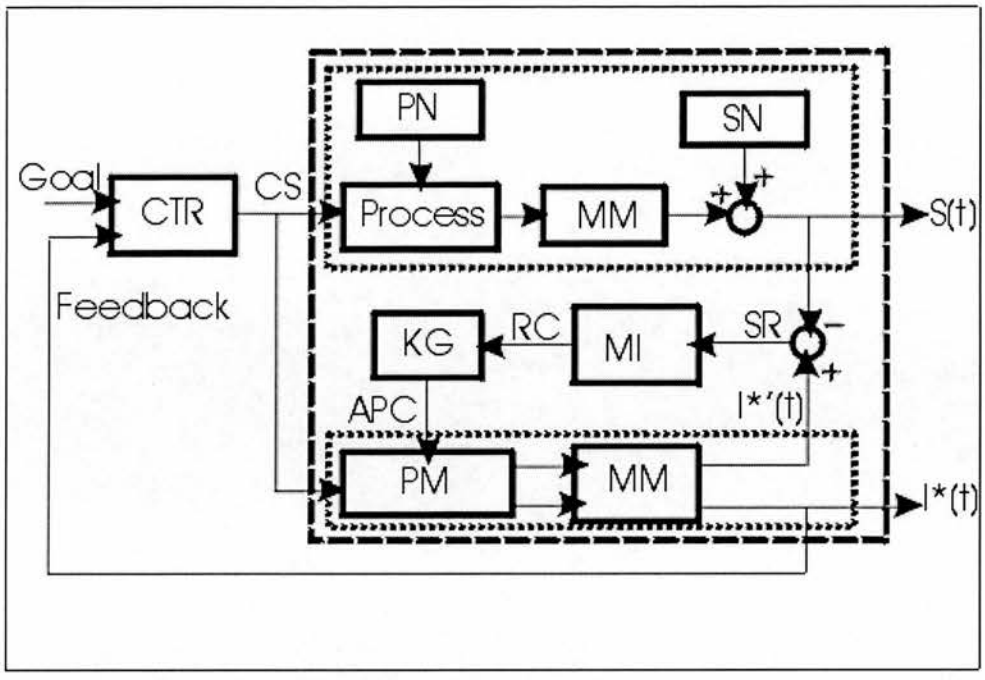


Figure 4.11 Control scheme blending pseudo-closed loop control and a Kalman filter (dashed rectangle), following Grush (2004). The upper dotted rectangle corresponds to the Controlled Object ("Plant"), the lower dotted rectangle the Emulator. The copy of the CTR output signal to the Emulator is referred to as the "efference copy" (c.f., figure 4.1 and 4.12). Abbreviations: controller (CTR), process noise (PN), measurement matrix (MM), sensor noise (SN), control signal (CS), measurement inverse (MI), Kalman gain (KG), process model (PM), sensory residual (SR), residual correction (RC), a posteriori correction (APC), *a priori* signal estimate I*'(t), actual observed signal S(t). I*(t) is the Kalman filter estimate of the observed S(t), but noise free (i.e., filtered). See text and Grush (2004) for more details.

Notice that as part of the measurement update, an *a priori* signal estimate *I*'(t)* is compared with the actual observed signal *S(t)* (figure 4.11), the difference being the "sensory residual" (Grush, 2004). Although the filter can alter the *a priori* estimate to completely eliminate the sensory residual, it usually does not do this because the residual is not entirely due to inaccuracy; in part it is due to sensor noise. The extent of the alteration is determined by the magnitude of the Kalman gain. At one extreme, if the Kalman gain is set to unity, the system comprises closed-loop control. On the other hand, if the Kalman gain is zero, none of the residual correction is applied, the *a priori* estimate is never adjusted on the basis of feedback, and the process model (emulator) evolves over time (acquires successive states) exclusively under the influence of its internal dynamic and the controller's signal (an estimate of which comprises the feedback sent to the controller) (Grush, 2004).

In cognitive processing terms, low Kalman gain corresponds to predominately "top-down" processing, whereas high Kalman gain corresponds to predominately "bottom-up" processing (with the filter adjusting the gain according to the level of uncertainty based on past experience) (Grush, 2004). In the case of imagery, Grush argues that the Kalman gain is zero (or close to zero), with the emulator's state evolving in time according to its own dynamic and controller efference copies, unaffected by sensory feedback correction. The motor command outputs are assumed to be suppressed (Grush, 2004). Of more relevance to psychiatric illness (and just as speculatively), if the Kalman gain were to be dysfunctionally fixed at zero, psychotic experiences such as hallucinations might result (see below).

Finally, Baev has described a similar but more complex model of hierarchical neural optimal control systems (NOCS) (figure 4.12). Although individual NOCS are based on a Bayesian framework, and adapt towards optimal functioning, they are not identical to a Kalman filter (Baev, personal communication). The method by which a hierarchy of stochastic control systems could adapt towards optimal functioning is not well understood theoretically. It has been observed that healthy self-regulating systems often exhibit fractal *1/f* scaling of the power spectra (where *f* is frequency), and measures of many diseases (Goldberger, Amaral, Hausdorff, *et al*, 2002; Peng, Hausdorff & Goldberger, 2000), and mood disorder (Woyshville,

Lackamp, Eisengart, *et al*, 1999), are associated with a marked loss of normal long-range correlations or "complexity". It has been speculated that long range correlations *in health* "may serve as a newly described organising principle for complex [hierarchical] non-linear [homeostatic] processes that generate fluctuations on a wide range of time scales" and "the lack of a characteristic scale [in health] may help prevent excessive mode-locking [driven periodic oscillations] that would restrict the functional responsiveness (plasticity) of the organism" (Peng, Hausdorff & Goldberger, 2000).

4.7 DISCUSSION

4.7.1 Summary and hypotheses

Control theory has been applied to the study of diverse physiological systems, and has also been used to describe many aspects of emotion, cognition and motor function. The neurological substrate of such functions includes the prefrontal BGTCLs: their normal function, and dysfunction in disease, may be consistent with the actions of negative feedback regulatory systems. Quantitative associative learning theories predict behavioural response to reinforcers, and are supported by extensive experimental evidence for brain predictive error signals. Many brain regions and functions have been proposed to operate as a Kalman filter. However, it can be argued that, whilst similar, the various theories are clearly not identical. Nevertheless, stochastic control theories can be viewed as a generalisation of deterministic theories, and various closed loop configurations can be accommodated within the same model. Clearly though, more work on the comparison of such theories is indicated. The study of depressive illness, described in chapter 5, used a Kalman filter model of brain function, as figure 4.11, within a pseudo-closed loop negative feedback control context. The study tested the hypothesis that, on the basis of numerous Kalman filter models of brain function, the brain would automatically create an optimal prediction of sensory input (winning or losing in a game), approximated by a Kalman filter, and that this would be compared to actual sensory input (feedback on actual winning or losing),

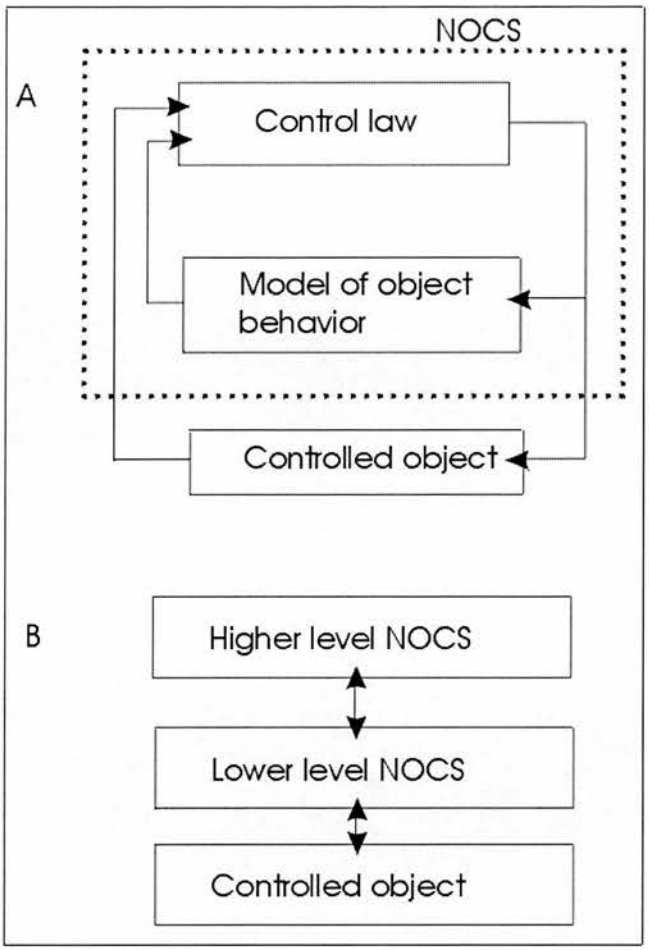


Figure 4.12 Baev's hierarchical stochastic model of brain function (Baev, 1998; Baev, Greene, Marciano, *et al*, 2002; Baev, Greene, Marciano, *et al*, 1995), which is based on extensive animal studies of "central pattern generators" . As in other theories, the actual afferent signal from the controlled object is compared with the predicted model signal, the difference forming an error signal. The "neural optimal control system" (NOCS) is enclosed within the dotted rectangle. The structure in "A" is repeated throughout the hierarchy in "B". This diagram is a considerable simplification of Baev's theory, which includes various control signals (not shown), and a quantitative formulation which allows tests of the theory. A stochastic basal ganglia thalamocortical model (Baev, Greene, Marciano, *et al*, 1995, figure 2) was also described and discussed in the context of various pathologies (see chapter 5).

to create a Kalman filter derived error signal. The fMRI data were then tested for the presence of this error signal. It was further hypothesised that depressed patients would systematically differ from healthy control subjects with regard to such an error signal (generate biased non-optimal predictions). The reasons for this second hypothesis are described in chapter 5.

The study described in chapter 6 investigated a number of reports of a subtle abnormality of structure in the brainstem of patients suffering from unipolar depressive illness. The authors of those reports speculated that their reported abnormality was consistent with a disruption of the monoaminergic fibre tracts in patients. As discussed earlier, there is extensive evidence for the monoaminergic systems exhibiting predictive error signals. A structural abnormality of these systems could therefore be hypothesised as disrupting the normal associative learning mechanisms. This might lead to impaired learning about reinforcers in the environment, and consequently be a basis for depressive symptoms such as anhedonia.

4.7.2 Future work

Chapters 5 and 6 describe tests of the above main hypotheses. Additionally however, there are broader implications of stochastic theories of brain function which are important to consider for future studies. In particular, it has been suggested that there is a growing consensus in neuroscience: "the brain learns and maintains an internal model of the external world" (Barlow, 1985, p37-46; Rao, 1999), and "the human brain generates an internal model to fit incoming information about the external world, and experiences the model rather than the information" (Picton & Stuss, 1994; Rao, 1999). For example, it has been suggested that the subjective experience of pain depends on active modelling of sensory input (Wall, 1993). "The brain constructs a virtual reality and fits this to the senses in a process that can be significantly altered by placebo. We experience the virtual rather than the real." (Picton & Stuss, 1994; Wall, 1993). Gray has proposed a different but potentially related theory, equating consciousness with the closed loop operation of the septo-hippocampal (figure 2.7; and possibly other) system(s) (Gray, 2004). In Grush's theory, raw sensory information constrains the configuration and evolution

of the (emulator based) representation: "perception is a controlled hallucination process" (Grush, 2004). Baev and colleagues additionally observe: "Long ago psychologists concluded that the brain creates models of the environment. The proposed theory [above] helps to understand these brain models" (Baev, Greene, Marciano, *et al*, 2002, p802)

In summary, there are a number of theories of subjective experience or consciousness which are based on the hypothesis that the brain actively models, rather than passively receives, incoming information. It has been claimed that consciousness corresponds with the model or emulation, not the basic information. Such theories may be compatible with theories of brain function based on stochastic feedback control: detailed discussion can be found elsewhere (Barlow, 1985; Picton & Stuss, 1994; Rao, 1999; Stuss, Picton & Alexander, 2001). For some related and very different theories see Gray (Gray, 2004). From the perspective of such theories, depressive illness may consist of dysfunctional models of the world and self, reflecting disordered associative learning, which psychotherapy attempts to alter. Such putative dysfunctional models could be manifested by negative automatic thoughts about self, surroundings and future; i.e., negative automatic thoughts become the "most likely" interpretation of the information (biased prediction). Although speculative, these theories offer a potential link between subjective experience (perhaps even including psychotic symptoms), "cognitive distortions" (as a form of biased automatic prediction), the effects of antidepressant medication (via predictive error signals and associated neural plasticity) and abnormalities of brain function reported in imaging studies of depressive illness. Importantly though, it is possible to test such theories: e.g., depressive illness should be associated with abnormal predictive error signals, when account is taken of localised specialisation of brain function (see chapter 5).

CHAPTER 5

NEURAL PREDICTIVE ERROR SIGNAL CORRELATES WITH DEPRESSIVE ILLNESS SEVERITY IN A GAME PARADIGM

5.1 INTRODUCTION

Reviews of imaging studies of depressive illness consistently suggest that various regions of the prefrontal and temporal lobes have abnormal function and structure (Drevets 2000; Ebmeier and Kronhaus 2002). Such regions include the anterior cingulate, orbitofrontal and dorsolateral cortex, hippocampus and amygdala. Whilst metabolic activity is frequently reported to be abnormal compared with controls, and structural abnormalities such as grey matter reduction have been reported, the mechanisms by which these abnormalities are related to, and perhaps cause the symptoms of depressive illness, are unknown. Elucidation of such mechanisms may allow development of more effective treatments for depressive illness.

Over the past decade there has been a rapid advance in understanding the neural mechanisms underlying associative learning (e.g. Schultz, Dayan & Montague, 1997; Schultz & Dickinson, 2000; Schultz, Tremblay & Hollerman, 1998). Associative learning allows humans and animals to predict the outcomes of important events. Learning occurs when the actual outcome differs from the predicted outcome, the difference between the two constituting a prediction error. There is considerable animal and human experimental evidence of neural activity in a number of brain regions representing such predictive errors (Schultz & Dickinson, 2000). Many of these regions are the same as those reported as having abnormal function and structure in depressive illness. I hypothesise that such abnormalities may be associated with a dysfunction of associative learning. Further, this may be reflected by different predictive error signals in depressed patients, when compared with healthy controls.

These are various reasons for such a hypothesis. Existing theories (Rolls, 1999; Solomon & Corbit, 1974) suggest that there is a direct link between reinforcers and normal emotions. Rolls argues that different emotions can be described and

classified according to whether the stimulus is pleasurable (positive) or aversive (negative), the intensity of the stimulus and the reinforcement schedule (Rolls 1999). For example, omission of (predicted) positive reinforcers leads to emotions such as frustration, through anger to rage, or sadness through grief to depression. Similarly, omission of (predicted) negative reinforcers involves emotions such as relief through pleasure to elation. By definition, the omission of an expected rewarding or aversive stimulus is associated with a predictive error signal. Consequently, Rolls' theory suggests that such an error signal arising in the context of unexpected reinforcing stimuli should be associated with two opponent groups of emotional responses depending on the nature of the stimulus. Depressed patients, who have a tendency to experience unpleasant emotions, would therefore be predicted to exhibit a different pattern of error signals from healthy subjects, who do not. A similar argument may be advanced on the basis of Solomon's opponent process theory (Solomon & Corbit, 1974).

Studies on animals have reported that the monoamine systems (e.g., dopamine and noradrenaline) exhibit predictive error signals, as does the acetylcholinergic system (Schultz and Dickinson 2000). It has long been recognised that virtually all empirically discovered antidepressants increase monoamine levels by a variety of different mechanisms. This observation led to the well known monoamine hypothesis of mood disorder (Bloom and Kupfer 1995; Schildkraut 1965), which simply states that depression is associated with a monoamine deficit, mania with an excess. An acetylcholine theory of mood disorder was also proposed (Bloom and Kupfer 1995). The limitations of these theories are well known (Owens, 1998). I suggest that antidepressants may exert their beneficial effects by direct actions on associative learning mechanisms, since the monoamine systems, and regions of the brain repeatedly implicated as being abnormal in studies of depressive illness, exhibit predictive error signals (chapters 2 & 4).

Various quantitative models have been developed to describe associative learning behaviour. These include the Rescorla-Wagner equation (Rescorla & Wagner, 1972) which is based on a linear prediction of reward associated with a stimulus, and the Temporal Difference equation (Sutton & Barto, 1998), which takes account of how an animal can predict the time of reward delivery within a trial.

Related learning models include the back-propagation rule and Pearce-Hall theory (Dayan, Kakade & Montague, 2000). The Kalman filter model also formalises the predictive relationship between stimuli and reinforcement, together with how the relationship is expected to change over time. Such expectation is related to the degree of uncertainty associated with each stimulus based on accuracy of prior predictions. At each time step, the future predictions are a compromise between prior information and recent observation (Dayan, Kakade & Montague, 2000). All these quantitative models contain a similar predictive error term (Dayan, Kakade & Montague, 2000).

Of particular relevance to this study, not only do these theories describe animal and human behaviour under situations of reward conditioning, but describe also the neural error signals from the prefrontal lobe and cerebellum discussed above. These seem to conform to the Rescorla-Wagner or Delta rule (Schultz & Dickinson, 2000). Dopamine neurones code an error in the prediction of reward which conforms to the Temporal Difference equation (Schultz & Dickinson, 2000). Control system models utilising Kalman filters have been proposed to account for a number of integrated functions in humans; e.g. the visual system (Rao, 1999), sensory adaptation (Grzywacz & de Juan, 2003), spatial orientation (Borah, Young & Curry, 1988), visuo-motor performance (Baddeley, Ingram & Miall, 2003; Borah, Young & Curry, 1988; Wolpert, Ghahramani & Jordan, 1995) and cerebellar function (Paulin, 1988). Further, a Kalman filter-like model of complex vestibular responses has been proposed (Glasauer & Merfield, 1997). Of particular relevance to the study of depressive illness, the hippocampus and the amygdala (plus ventral striatum and dopaminergic system) may operate together as a Kalman filter (Dayan, Kakade & Montague, 2000). On other evidence, it has been suggested that the hippocampus may operate as a Kalman filter (Bosquet, Balakrishnon & Honavar, 1999).

Recent imaging studies of healthy subjects have investigated predictive error signals. Berns described a predictive error signal in the ventral striatum and orbitofrontal cortex using an fMRI study (Berns GS et al., 2001). In another study, unexpectedly late reward delivery was found to activate the ventral striatum (Pagnoni G et al., 2002). Electrophysiological imaging has identified a behavioural error signal localised to the anterior cingulate, although this may reflect error monitoring

unrelated to motivation (Dehaene S et al., 1994). A study of prediction error associated with aversive conditioning (thermal pain) reported cerebellar activation (Ploghaus A et al., 2000). Additionally, a temporal difference error signal was found in the orbitofrontal cortex, ventral striatum and cerebellum of subjects undergoing reward conditioning (O'Doherty JP et al., 2003). In another recent study, a passive learning task was used to demonstrate temporal prediction error signals in the striatum (McClure, Berns & Montague, 2003).

Given extensive evidence of predictive error signals occurring in brain regions reported as abnormal in many studies of depressive illness, and of a direct link between emotion and reinforcers, it was hypothesised that depressed patients would differ from healthy controls with regard to a predictive error signal arising as a consequence of experiencing unexpected rewarding or aversive events. This signal was expected to arise from limbic and paralimbic brain regions known to be associated with processing these stimuli: e.g. orbitofrontal cortex, rostral medial prefrontal cortex, ventral striatum and hippocampus. Not only was the signal expected to differ within localised regions, but also to differ between such regions, reflecting alteration in the influence of one neuronal system on another. The approach of using computational models to generate regressors for neuroimaging studies appears promising. Therefore, we hypothesised that the error signal could be modelled by the difference between the actual observed stimulus and a Kalman filter prediction.

5.2 MATERIALS AND METHODS

5.2.1 Subjects

Permission for the study was obtained from the local ethics committee and written informed consent obtained from each subject. A total of 30 right-handed subjects participated in the study. One subject was unable to tolerate scanning resulting in data from 15 patients and 14 controls. Subjects were excluded if they had a history of significant head injury or structural brain abnormality (including vascular disease), substance misuse, physical disorder known to affect brain metabolism (e.g., epilepsy) or receiving medication which might alter brain

	Patients	Controls	p
Age	45.9	43.0	n.s.
Females	11	7	n.s.
NART	12.4	8.5	n.s.
BDI	36.9	1.1	<0.001
Hamilton	27.3	-	-

Table 5.1 Mean clinical characteristics of patient and control groups; difference not significant (n.s.)

	Total daily dose / mg
Sertraline	50
Fluoxetine	60
Phenelzine	45, 75
Lithium carbonate	1000, 700, 1000, 800
l-tryptophan	3000
Venlafaxine	375, 150
Mirtazepine	15, 45
Moclobemide	600, 600, 600
Citalopram	20, 20
Imipramine	150
Amitryptiline	50
Pindolol	15
Quetiapine	300
Sodium valproate	200
Diazepam	5
Zopiclone	7.5
No medication	One patient

Table 5.2 Total daily dose of each medication for each patient.

metabolism (e.g., steroids). Subjects were additionally excluded if they had deliberate (e.g., cardiac pacemaker) or accidental implanted metal. Subjects had to be without serious physical disease. The upper age limit was not limited to 65 years. All subjects completed a BDI (Beck, Ward & Mendelson, 1961) to allow a simple comparison self-rated depressive symptoms between patient and control groups. The BDI was not used for any other purpose.

Patients were recruited from inpatients and outpatients at the Royal Edinburgh Hospital. For inclusion they had to have an unequivocal single diagnosis of unipolar depressive illness in the opinion of both the referring consultant and the author after an interview and review of their notes. The severity of depression had to exceed a score of 20 on the Hamilton depression sale. If there was a history of any other diagnosis (e.g., alcohol misuse, borderline personality disorder etc.), or a possible previous episode of elevated mood, the patient was excluded. Since it was so important to exclude patients with bipolar disorder (see chapter 6), potential subjects admitting to a family history of bipolar disorder were excluded. Preference was given to patients with the most severe illness who appeared able to take part, since previous imaging studies have suggested that the most marked abnormalities are to be found in such patients. Preference was also given to identifying patients who had the most treatment resistant illness, since they were very likely to be ill at the time of scanning, despite frequent difficulties in booking scanner times. No patients were believed to be actively psychotic at the time of scanning though two had a history of psychotic unipolar illness. Study planning included determining whether the paradigm was practical for patients with psychomotor retardation by use of a scanner simulator, examining timing data logs of button presses, and discussion with the subjects. This indicated that the paradigm appeared suitable and so such patients were not excluded. Patients with significant suicidal intent were excluded despite the fact that they constitute a sizable proportion of the population of depressed patients. The reason was partly because of the difficulties in safely transporting such patients to and from the scanner, and partly because of other difficulties. For example, one patient who was approached expressed the belief that medical scanners were dangerous and likely to malfunction causing the death of the person being scanned. This was his reason for wanting to take part, which clearly

called into question the validity of his potential consent. Another patient who was considered took a large overdose of medication the day before she was due to be scanned. Patients with significant comorbid anxiety (e.g. panic attacks) were not excluded, though those admitting to having a specific problem with claustrophobia were excluded due to the nature of the scanner environment. All patients but one were receiving a variety of psychotropic medication (table 2) which had remained at constant dose for at least two weeks prior to scanning. There is no accepted way to calculate equivalent doses of different medications, or how to take account of treatment combinations (e.g., lithium augmentation of antidepressants). However, to attempt to explore a possible relationship between total amount of medication and severity of illness, the dose of each medication was calculated as a percentage of the maximum prescribable dose. In the case of patients receiving more than one medication these were combined in two different ways: as an average and as a total. Regression of each against Hamilton score was not significant ($p=0.84$ and $p=0.35$ respectively).

Controls were acquired from work colleagues, friends and relatives. Details of the subjects are given in table 1. The control group was matched on the basis of age, NART (Nelson & Wilson, 1991) and percentage of female subjects. Potential controls were warned that they would be asked about symptoms of psychiatric illness, alcohol use, and general medical history, and only if they were comfortable with this, were they further assessed to determine whether they might be suitable. A brief interview was done to determine whether any admitted to current symptoms of depressive illness or were misusing alcohol or other substances. If so, they were excluded. They were asked whether they had ever been diagnosed in the past as suffering from a depressive (or other psychiatric) illness, or advised to reduce alcohol consumption, and if they had ever been prescribed antidepressants. If so, they were excluded. Since it was so important to exclude patients with a bipolar depressive illness, potential controls were asked about such a family history, and would have been excluded if they admitted to this. As indicated above, potential controls were also asked about any previous head injury, any disease affecting the brain, and current medication. The medical case notes of the control subjects were not examined.

There are various limitations in the above method of identifying patients. For example, structured interview methods such as the Schedule for Affective Disorders and Schizophrenia (SADS) (Overall & Gorham, 1962), Structured Clinical Interview for DSM (SCID) or Present State Examination (PSE) (Wing, Cooper & Sartorius, 1974) were not done and it could be argued that using these methods would have improved the reliability of diagnosis over standard clinical approaches. However, these methods have various disadvantages: SADS can take up to two hours to complete and the SCID is rarely used which limits its interpretation and comparison with other studies. The PSE is time consuming though good for the assessment of psychotic disorders: unfortunately it is less useful for characterising affective disorders. These limitations should be considered in the context of the available resources for this study. The two primary limitations were: a) funding for a maximum 30 subjects, and b) all data had to be collected and analyzed in its entirety by the author (in the context of various other commitments and a limited period to complete the study). Consequently, there was no attempt to acquire a large sample of patients. The advantage was that preference could be given to patients for whom a single diagnosis of unipolar depression was in least doubt. Larger studies aiming to include far more patients would probably need to use structured interview methods since they would necessarily consider patients for whom the diagnosis was less certain.

Additionally, it is necessary to mention that the female menstrual cycle has effects on mood, which may be related to altered levels of many biological variables. However, the percentage of females in both the patient and control groups were deliberately balanced. Additionally, the only correlate of interest was a measure of mood (Hamilton score) which presumably would have captured any change in mood due to the stage of the menstrual cycle. Furthermore, there are no previous reports of an effect of the menstrual cycle on neural predictive error signals. Consequently, it seems unlikely that the findings reported later could be attributable to the menstrual cycle.

There are additional limitations with regard to the way that controls were selected. Significant abnormality in the patient group is determined as much by the selected controls as by the patients. The above matching allowed acquisition of

controls at a rate which was practical for successful completion of the study. However, it is likely that controls and patients may have differed on a number of other factors such as socio-economic status and adversity. Not only are there problems with regard to measurement of such factors, but it is unclear whether such factors should be controlled. For example, it is now believed by some that early structural scanning studies of patients with schizophrenia, employing case-control designs, may have had a tendency to recruit "supernormal" controls, finding factors more related to the severity of illness than causation. Similar problems could have confounded this study. Alternatively, there are suggestions that adversity is associated with structural hippocampal changes and an increased risk of depressive illness. Consequently, such adversity could be an important influence on the mechanism by which patients develop depressive illness. Although such possible bias could have been reduced by random sampling of the population of controls this is unlikely to have been practical, since it would have extended the duration of the study. Additionally though, the necessary exclusion and inclusion criteria would itself have resulted in non-random sampling. Recognition of such problems has resulted in arguments for "pragmatic studies" with very limited exclusion criteria, since such studies include subjects more representative of the patient population of interest. A pragmatic study was not used here since depressive disorder and not alcohol misuse (etc.) was the primary concern.

5.2.2 Imaging

Gradient ecoplanar T_2^* weighted BOLD contrast images were acquired on a GE Medical Systems Signa 1.5T MRI scanner. For each subject, a total of 30 axially orientated (parallel to AC-PC line) contiguous 5mm thick interleaved slices were obtained for each image volume, 500 volumes being acquired with a TR of 2.5 ms. The TE was 40 ms, flip angle 90°, FOV 24 and matrix 64*64. The first 4 such volumes were discarded to allow for transient effects. A T_1 weighted image was obtained as a record of detailed anatomical structure with 1.7 mm thick contiguous slices, matrix 256*192 and FOV 22. Additionally, a routine clinical T_2 weighted image was obtained with 5 mm thick slices and 1.5 mm inter-slice gap, matrix

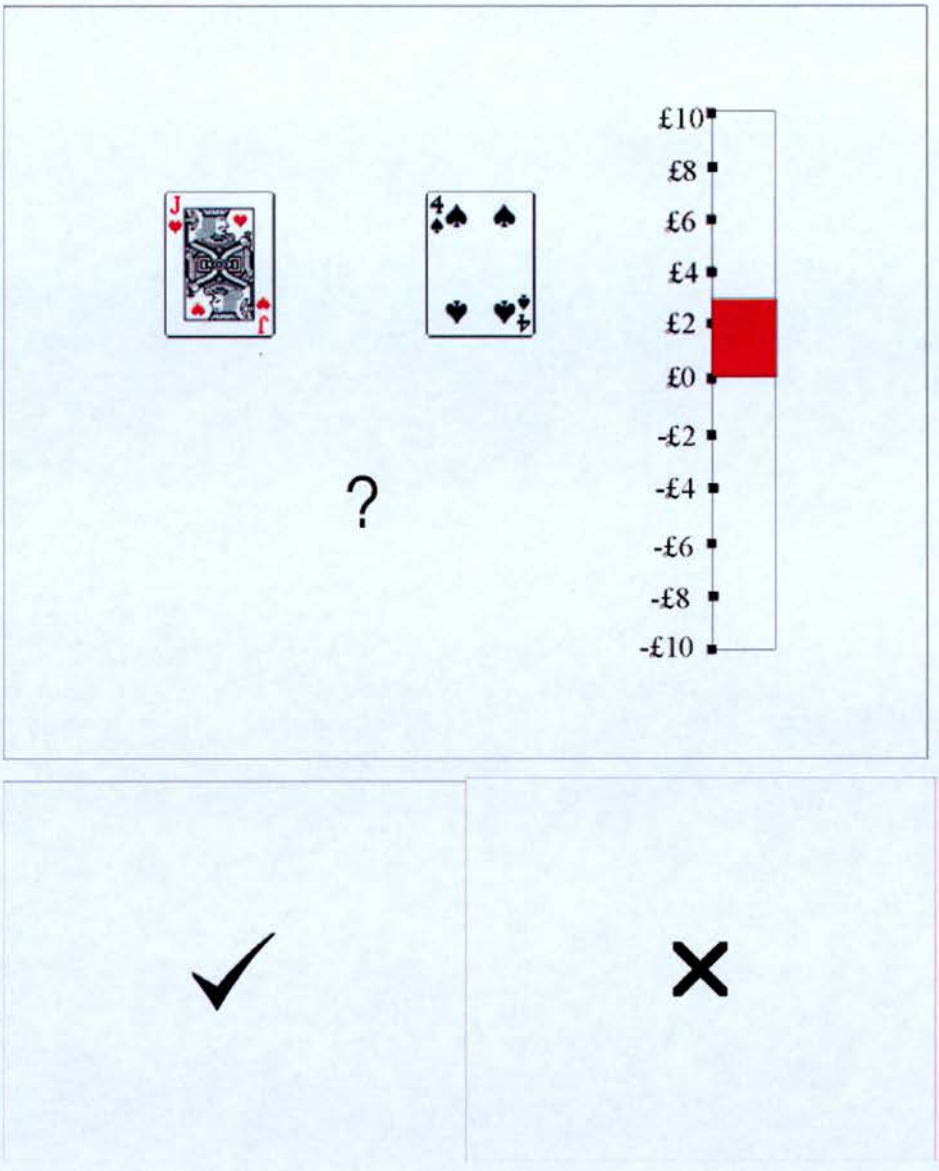


Figure 5.1 Displays observed by subjects during scanning. Subjects were presented with the card choice screen followed by a screen with either a tick or cross.

256*256 with FOV 24 and reported by a neuroradiologist for signs of neurological disease.

5.2.3 Paradigm

The paradigm which was used was chosen primarily because it appeared practical for patients to perform in an MRI scanner, recognising that some would have a moderate or severe depressive illness. It was also chosen because the task had been shown to be associated with brain activation in specific regions of interest in three independent studies on healthy controls using different image processing strategies (Akitsuki, Sugiura, Watanabe, *et al*, 2003; Critchley, Elliott, Mathias, *et al*, 2000; Elliott, Friston & Dolan, 2000). Subjects were presented with pairs of playing cards, one red and one black, with the red card always on the left; figure 5.1. In study planning several patients and controls took part in a scanner simulator. Some patients were initially quite reluctant to take part, feeling that they would be unable to perform the task after listening to a description. It was therefore found useful, and in some cases necessary, to suggest a plausible approach to playing the game; "rules" which if guessed correctly would result in winning. Therefore, in the actual study, all subjects were told that one card was "correct" according to a pre-programmed rule; e.g., "it might be the red card or the black, or the higher or the lower card". The rule would stay the same for some time then change and subjects were to try to work out the correct rule by guessing. Subjects had 2.5 sec to guess. They were told that if they did not guess in time a random guess would be made for them. Following the 2.5 sec card presentation, feedback on whether a correct card had been chosen was provided for 1 sec by a screen with either a tick or cross. Each correct choice resulted in winning £1, each incorrect choice with losing £1. The running total was displayed on a bar on the right hand side of the card choice screen. Subjects were told they were not playing for actual money, but for "points" and the objective was to try to end the game with the maximum number of points.

All subjects were allowed a brief practice session on a computer before undertaking the task in the scanner. An important feature of the paradigm was that whilst the subjects were led to believe that their responses determined the outcome,

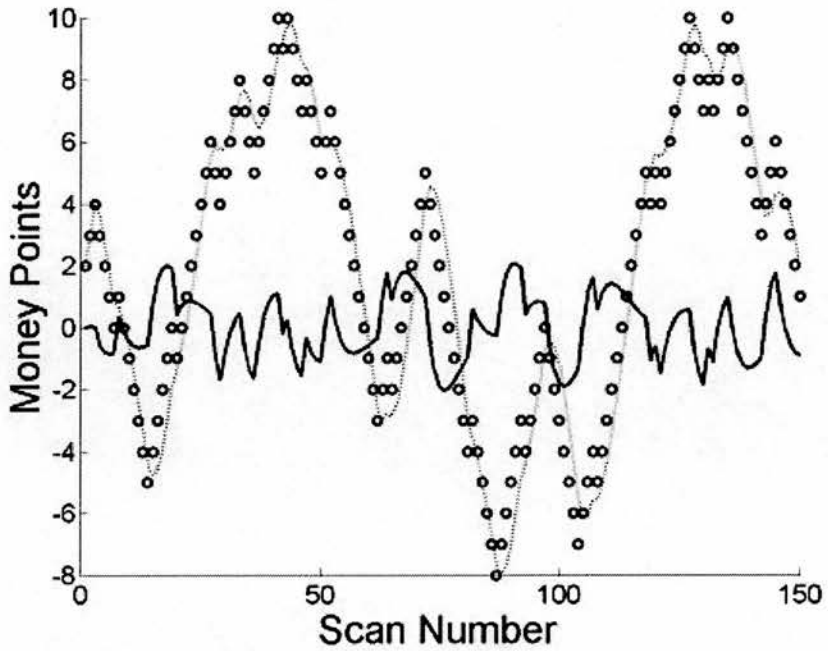


Figure 5.2 Sequence of money bar heights (o), Kalman predicted height (...) and prediction error (___). Only part of the total scan period is shown for clarity.

the feedback sequence was entirely predetermined; figure 5.2. The only variables of relevance are the immediate winning or losing and the trend in such feedback. The task merely provides a plausible context for winning and losing independent of actual performance. This ensured that each subject received identical reinforcement stimuli independent of actual ability which was expected to be relatively impaired in the patient group. Unipolar depressed patients have been reported to have widespread impairments on almost all CANTAB neuropsychological tests, this impairment correlating with depressive illness severity (Elliott, Sahakian, McKay, *et al*, 1996) and involving dysfunction of ventromedial prefrontal brain regions implicated in associative learning (Elliott, Sahakian, Michael, *et al*, 1998).

The paradigm was implemented on the IFIS fMRI system (Psychology Software Tools; http://www.pstnet.com) and a log file acquired for each subject which included precise recording of "tick or cross" feedback presentation time in relation to scanner radio-frequency pulses. Pairs of cards were presented 12 times, this being followed by a 5 sec rest period, the whole block being repeated 24 times. This resulted in a total scanning period of 22 min (including a prompting message to subjects at the beginning of each block). Discussion with subjects indicated that all were doing their best to win, and data logs indicated that no subject passively observed the screen without making card choices. No subject guessed that the sequence was predetermined, which was considered crucial for "emotional involvement" in the task.

5.2.4 Predictive Error Signal

A new money bar height (x_1,x_2) at time t was defined as the old height at presentation time (t-1), plus rate of change of bar height (dx_1,dx_2), plus noise (w). According to Kalman filter theory (Kalman, 1960) this can be expressed:

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ dx_1(t) \\ dx_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \\ dx_1(t-1) \\ dx_2(t-1) \end{bmatrix} + \begin{bmatrix} w_{x1} \\ w_{x2} \\ w_{dx1} \\ w_{dx2} \end{bmatrix} \qquad 5.1$$

Since only the height of the bar at a given presentation time is observed:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} x_1(t) \\ x_2(t) \\ dx_1(t) \\ dx_2(t) \end{bmatrix} + \begin{bmatrix} v_{x1} \\ v_{x2} \end{bmatrix} \qquad 5.2$$

The above calculation was implemented by modifying Murphy's existing Matlab (Mathworks Inc) script, which made use of his associated Kalman filter Toolbox (http://www.ai.mit.edu/~murphyk/Software/Kalman/kalman.html). The error signal was calculated as the difference between the observed and predicted money bar heights; figure 5.2.

Whilst the task apparently involves choosing between two cards, it is not necessary to model this aspect since the card sequence is random and carries no information for prediction. Also, the above equations model two variables; the height of the money bar and the time of stimulus presentation. The advantage was that in this form the equations are analogous to a two-dimensional tracking task, a classical application of Kalman filtering in engineering control systems (which facilitated checks that the calculations were correct). The variations in stimulus presentation time were very small, and did not contain any useful information for predicting bar height, so were subsequently disregarded.

5.2.5 Basic Image Analysis

SPM99 (http://www.fil.ion.bpmf.ac.uk/spm/) was used to process the images. Once the processing sequence was finalised using the standard SPM99 interface it was automated using custom Matlab routines, which made use of the KULeuven interface functions (http://www.kuleuven.ac.be/radiology/Research/fMRI/kulSPM/index.html). Images were first corrected for differences in image acquisition time between slices and then the time series were realigned to the first scan in each series. Spatial normalisation to the SPM99 echoplanar template (which conforms to Montreal Neurological

Institute (MNI) space) followed with images being subsequently smoothed with a 6 mm isotropic gaussian kernel.

To determine whether an error signal was present, an exploratory fixed effects analysis was calculated for the control group with data for each subject entered as different "runs" of the same subject. An event related non-stochastic design was used with variable onset stimuli defined by the IFIS log file data entries for "tick or cross" feedback presentation time. The neural response was modelled as a linear function of the Kalman derived error signal convolved with the basic SPM99 haemodynamic response function. Scaling was used for global normalisation and high pass filtering was done. Significant regions were only reported for $p<0.05$ after correction for the whole brain volume using the false discovery rate (FDR) method as implemented for SPM99. The patient group was similarly analysed.

To directly compare both groups and allow generalisation to the patient population of interest, a random effects analysis was also done. First level model specification comprised a fixed effects analysis calculated as above but for each subject separately. The contrast images from the first level were entered into two second level analyses. The first was an independent t-test to directly compare patient and control groups. The second comprised a test of correlation with Hamilton depression rating for the patient group to test for systematic changes in error signal with depression severity. Results were only reported for $p<0.05$ after correction for multiple testing using the conservative family-wise error rate (FWE) method.

Specific regions of interest were defined *a priori* and justified use of a 10 mm diameter SPM99 small volume correction. This was chosen for consistency as being the same as used for the connectivity analysis (since the same regions appeared in both analyses). Additionally, it was only slightly larger than the filter used for smoothing which may assist signal detection.

5.2.6 Selection of Regions of Interest

The regions about which there were *a priori* hypotheses are as defined in table 3. These were selected on the basis of being associated with emotional processing in numerous previous studies on healthy subjects and abnormal activity in many studies on patients with depressive illness. As stated above, the experimental

paradigm used here ("Elliott's paradigm") was selected partly on the basis of activating similar regions in three previous studies on healthy subjects. It was recognised that it may not be possible to examine some of the structures due to signal loss resulting from the susceptibility artefact (Lipschutz, Friston, Ashburner, *et al*, 2001). Such structures were later identified from inspection of the SPM99 mask file.

The rostral medial prefrontal cortex was reported active in all three previous studies using Elliott's paradigm although the precise location varied between studies. The center of the volume chosen was taken from a previous stereotactic meta-analysis of segregation of emotional and cognitive function in the prefrontal cortex (Steele & Lawrie, 2004a). This meta-analysis was based on a decade of published studies on healthy subjects, the motivation for that study being justification of *a priori* defined regions of interest in future studies. The lateral OFC was reported active in two of the three previous studies using Elliott's paradigm; again the precise location varied between studies. Therefore, the center of the region of interest was taken from the meta-analysis.

The hippocampus and parahippocampal gyrus were reported active in one of the three previous studies and a number of loci were reported. This structure is of particular interest due to previous reports of abnormal activity in depressive illness. Since a similar meta-analysis of segregation of temporal lobe function was not available, the regions of interest were selected on the basis of known anatomy. The parahippocampal gyrus was of particular relevance since previous studies have localised a comparator error signal to this structure rather than the hippocampus itself (Bosquet, Balakrishnon & Honavar, 1999; Gray & McNaughton, 2000). Since there may be segregation of function along the antero-posterior axis (Strange, Fletcher, Henson, *et al*, 1999) the parahippocampal gyrus was divided into two regions, one anterior and the other posterior. The WFU_PickAtlas (Maldjian, Laurienti, Kraft, *et al*, 2003) Matlab program was used to visualise the anatomy and define two centers of regions of interest for the parahippocampal gyrus and an entire region of interest for the hippocampus.

	Center of region of interest
Rostral anterior cingulate	(±5,46,18)
Lateral orbitofrontal cortex	(±42,28,-16)
Ventral striatum	(±10,10,-6)
Anterior parahippocampal gyrus	(±26,-9,-21)
Posterior parahippocampal gyrus	(±32,-46,-10)
Hippocampus	vox
Anterior cerebellum	vox
Posterior cerebellum	vox

Table 5.3 Predefined regions of interest. Coordinates correspond to the center of a 10 mm diameter volume in MNI space. Regions defined by a group of voxels selected using the WFU_pickatlas program are indicated ("vox").

The ventral striatum was reported active in two previous studies although rather different loci were reported. This is a relatively large and well defined structure (Mai, Assheuer & Paxinos, 1998) which can be visualised on a high resolution T_1 weighted MRI scan which conforms to MNI space (http://www.mrc-cbu.cam.ac.uk/Imaging/mnispace.html). The coordinate stated in table 3 was estimated as being the center of this structure on the high resolution scan. However, imaging of the ventral striatum is known to be affected by the susceptibility artefact.

Two of the three previous studies reported cerebellar activation although with considerable variation in reported location. Predictive error signals have long been known to occur in this structure. There is increasing evidence of the cerebellum being associated with emotional and cognitive function (Schmahmann, 1998) in addition to established motor function. Although segregation of function in the cerebellum is much less well established compared with the basal ganglia, most evidence probably implicates the anterior lobe in emotional processing and autonomic function (Heath, 1964). Therefore, two regions of interest were defined in the cerebellum; one in the anterior lobe and the other in the posterior lobe. Again, the WFU_PickAtlas program was used to visualise the anatomy and define the regions of interest.

5.2.7 Connectivity Analysis

The SPM99 volume of interest (VOI) routine was used to extract a time series for each subject for a 10 mm spherical volume centred at the coordinates shown in table 3. A custom Matlab routine was written to automate this extraction procedure and calculate the mean time series for each region. From this was calculated a covariance matrix for each subject. The anatomical model used for the effective connectivity analysis is shown in figure 5.3. This model was based on well established anatomical connections between the above regions of interest; specifically, the re-entrant basal-ganglia thalamocortical loops (Alexander, Crutcher & DeLong, 1990) and the re-entrant cortical-pontine-cerebellar-midbrain-thalamic

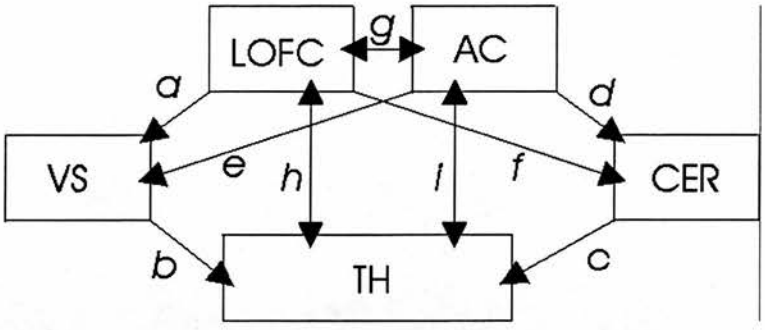


Figure 5.3 Anatomical model used for structural equation modelling. Lateral orbitofrontal cortex (LOFC), rostral anterior cingulate (AC), ventral striatum (VS), thalamic dorsomedial nucleus (TH), anterior cerebellum (CER). Paths *a-i* were estimated.

loops (Schmahmann, 1998). In both cases there is evidence of the loops being anatomically segregated, although the evidence is most complete for the basal ganglia.

Whilst it is necessary to pre-specify subregions of functionally segregated extended cortical structures for such modelling, the situation is different for small subcortical structures. In the latter case, structures which are too small to be clearly resolved have to be omitted from the anatomical model resulting in considerable simplification of the investigated connections. For example, the ventral pallidum has been omitted from the basal ganglia loop and the brainstem structures from the cerebellar loop. The hippocampus has been omitted from the anatomical model shown in figure 5.3 due to limited connections with the other regions of interest. The strongest anatomical connections are with the medial orbitofrontal cortex, caudal anterior cingulate and dorsolateral prefrontal cortex (Steele & Lawrie, 2004b). Since it was difficult to justify inclusion in the effective connectivity analysis, a separate functional connectivity analysis was used to explore changes in hippocampal connectivity to other regions of interest. To complete the loops it was necessary to define an additional region of interest in the thalamus (±10,-20,9). This was done using the high resolution T_1 weighted MRI scan.

Structural equation modelling was used for effective connectivity analysis. The method used was the Reticular Action Model (RAM) (McArdle & McDonald, 1984). Such a model is defined by three matrices defining asymmetric (A) connections, symmetric (S) connections, and a filter (F) matrix which relates observed to latent variables. On the basis of figure 5.3, these are defined as (Neale, 1997)

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a & e & 0 & 0 & 0 \\ 0 & 0 & b & 0 & c \\ f & d & 0 & 0 & 0 \end{bmatrix} \qquad 5.3$$

$$S = \begin{bmatrix} j & g & 0 & h & 0 \\ g & k & 0 & i & 0 \\ 0 & 0 & l & 0 & 0 \\ h & i & 0 & m & 0 \\ 0 & 0 & 0 & 0 & n \end{bmatrix} \qquad 5.4$$

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad 5.5$$

where *a-i* are path values to be estimated and *j-n* are the residual variances for each of the observed regions of interest. Given particular path values, a predicted covariance matrix (P_m) was calculated from (Neale, 1997)

$$P_m = F * (I - A)^{-1} * S * \left((I - A)^{-1}\right)' * F' \qquad 5.6$$

Optimal path estimates were obtained by minimising (Tabachnick & Fidell, 1996)

$$R = df * tr\left((O_m - P_m)^2\right) \qquad 5.7$$

where *tr* is the trace of the matrix, O_m is the observed covariance matrix and *df* is the degrees of freedom. The goodness of fit between the model and observed data was assessed using the standardised root mean square residual (*SRMR*) (Tabachnick & Fidell, 1996)

$$SRMR = \left[2 * \sum_{i=1}^{q} \sum_{j=1}^{i} \left(s_{ij} - \sigma_{ij}\right) \Big/ q(q+1) \right]^{1/2} \qquad 5.8$$

where s and σ are elements of P_m and O_m respectively and *SRMR* takes values between 0 and 1 with *SRMR*<0.05 indicating a particularly good fit. The above

calculations were implemented in Matlab using custom routines which made use of Adaptive Simulated Annealing (ASA) (Ingber, 1989) (http://www.ingber.com) for function minimisation. The ASA C code was interfaced to Matlab by use of the ASAMIN "mex" C code (http://www.econ.ubc.ca/ssakata/public_html). ASA was used in preference to simpler minimisation routines since it is generally recognised that identification of the global function minimum in structural equation modelling is made difficult by the frequent presence of local minima.

For model identification a necessary but insufficient requirement is that the number of estimated parameters be less than the number of observations. Therefore, following previous similar modelling (Grafton, Sutton, Couldwell, *et al*, 1994), the residual variance *g-n* was fixed. As a test for under-identification the ASA minimisation for each dataset was re-run 50 times with random starting values. Data were standardized to allow calculation of the *SRMR*. When this was done it was found that the likely global *F* minimum was obtained in over 98% of re-runs. Very large values of standard error (*SE*) can indicate problems with model identification (Neale, 1997) and therefore the *SE* associated with each parameter estimate was calculated on each re-run from

$$SE = \left[diag(H)^{-1} \right]^{1/2} \qquad 5.9$$

where *H* is the hessian matrix for minimum *R* and *diag* is the diagonal of the matrix. Such *SE* values were always of plausible magnitude. More importantly, the path values were always unique for a given value of *R*, consistent with the model being identified (Neale, 1997).

Empirically, it was found that setting *g-n* to 1.0 was associated with a minimum *SRMR* across data sets. This is also referred to as standardisation (Lawley & Maxwell, 1971). The *SRMR* was calculated for the 58 data sets (29 subjects, both hemispheres) and found to form a positively skewed distribution with a median of 0.06. The above method therefore achieves an acceptably good fit between the model and data. The path values obtained for each subject were regarded as summary statistics from a first level analysis and entered into subsequent second level linear regression analyses. This comprised a random effects design. No

correction was made for multiple testing because the paths were defined *a priori* and only a general pattern of change was of interest.

5.2.8 Susceptibility Artefact

The susceptibility artefact is known to cause signal loss and displacement in various regions; e.g., orbitofrontal cortex, anterior cingulate, ventral striatum, anterior brainstem and inferior temporal lobe (Lipschutz, Friston, Ashburner, *et al*, 2001). SPM99 automatically calculates a binary mask, which only includes voxels with a signal intensity at least 80% higher than global mean. Figure 5.4 shows a typical binary mask from a first level analysis on a *single* subject overlaid on to the high resolution T_1 scan (described above) using MRIcro (http://www.cla.sc.edu/psyc/faculty/rorden/). Some signal loss is clearly present in these regions, particularly the orbitofrontal cortex and subgenual anterior cingulate.

A fixed effects analysis or a second level of a random effects analysis in SPM99, both calculated for a group of subjects, generates a binary mask, which only includes voxels with signal present in all subjects. Since there are variations in the location of the signal loss in individual subjects, the effects of any signal loss is magnified. Figure 5.4 shows the binary mask from a second level analysis with substantial signal loss in many regions and consequent inability to detect task related activation or deactivation. This image defines the limitations of such *voxel based analyses* and indicates that the medial and some of the lateral orbitofrontal cortex plus ventral striatum could not be properly investigated. Additionally, much of the superior cerebellum could not be investigated; however, in this case the signal loss is confined to being very close to the midline so more lateral regions were not excluded. There is no prospect of examining subgenual anterior cingulate and brainstem, which although not included as one of the pre-specified regions of interest here, are implicated as having abnormal function in depressive illness. It was possible to examine the rostral anterior cingulate, hippocampus and to a limited extent the cerebellum and the lateral orbitofrontal cortex.

The effect of the signal loss on a *region of interest analysis* in which the signal from an extended region is averaged (e.g., in the connectivity analyses) may be less (if limited signal loss occurs in only part of the region) as illustrated by figure

5.4 (top). Such an approach allowed the lateral orbitofrontal cortex, ventral striatum and cerebellum to be investigated in the connectivity analyses. There are a number of alterations to image acquisition procedures which can be adopted to minimise susceptibility dependent signal loss (Lipschutz, Friston, Ashburner, *et al*, 2001). Unfortunately, these have not been implemented on the Edinburgh scanner.

5.3 RESULTS

5.3.1 Voxel Based Analyses

Fixed Effects Analyses Figure 5.5 shows the results of the fixed effects analysis for the control group. As expected, there are widespread regions of brain activation and deactivation corresponding to the Kalman derived error signal. It should be noted that these analyses only allow inferences about the particular subjects that took part in the study and qualitative comparison between the two groups is not statistically justified.

Random Effects Analyses The random effects design comparing the patient and control groups allowed generalisation to the patient and control populations. It showed enhanced signal in the rostral anterior cingulate and parahippocampal gyrus of the patient group (table 4). There were no significant decreases in brain activity between groups. Considering the patient group alone, multiple brain regions correlated positively with the Hamilton depression rating of illness severity when thresholded at $p<0.001$ uncorrected. After allowing for multiple testing incorporating a small volume correction, significant correlations remained ($p<0.05$) in the rostral anterior cingulate and posterior parahippocampal gyrus (figures 5.6-5.7, table 5.4).

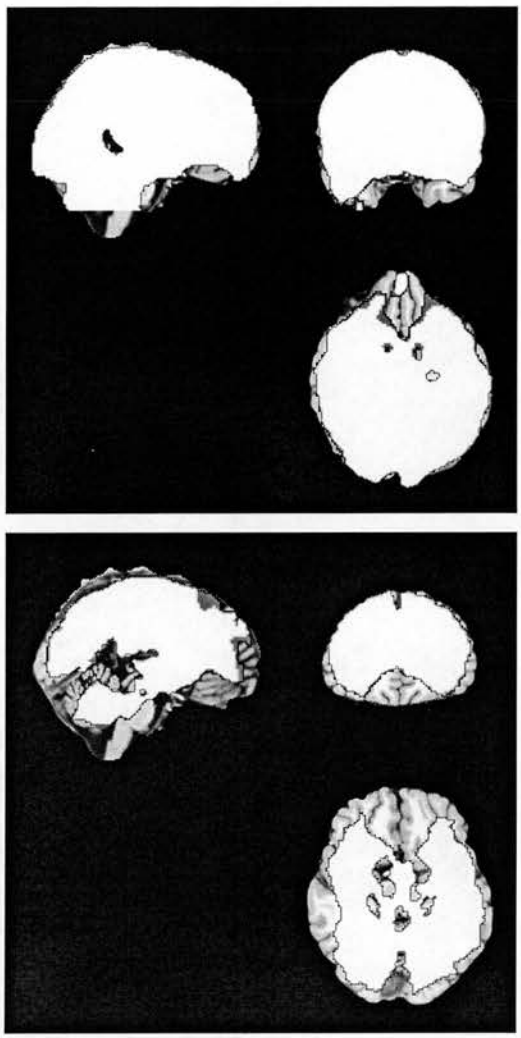


Figure 5.4 White regions are included in the SPM99 calculations. A mask file for a typical first level analysis of a single subject is shown with signal loss mostly in the orbitofrontal cortex and anterior brainstem (top). Also shown is the mask file calculated for the second level of the voxel based random effects analysis indicating regions where image data is available for all subjects at the first level (bottom).

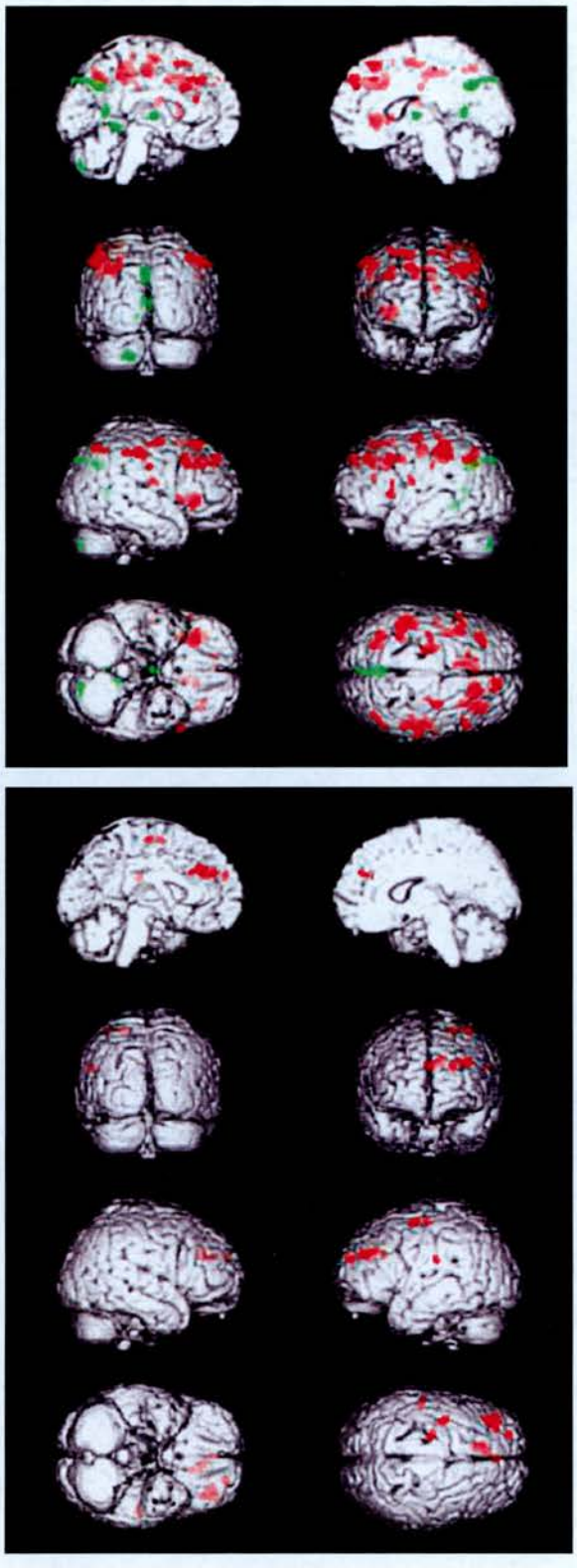


Figure 5.5 Fixed effects analyses showing regions of deactivation (red) and activation (green) for controls (top) and patients (bottom). Results are significant ($p<0.05$) after correction for whole brain volume.

		z	p	MNI
Rostral anterior cingulate	+G	3.41	0.039	(0,46,18)
Rostral anterior cingulate	+H	3.77	0.020	(6,40,6)
Parahippocampal gyrus	+G	3.41	0.019	(-32,-54,-4)
Parahippocampal gyrus	+H	3.51	0.014	(-34,-40,-16)
Anterior cerebellum	-H	4.15	0.046	(14,-40,-16)

Table 5.4 Voxel based random effects analysis with SPM99 small volume correction; increased signal in patients compared with control group (+G), positive correlation (+H) and negative (-H) with Hamilton depression rating.

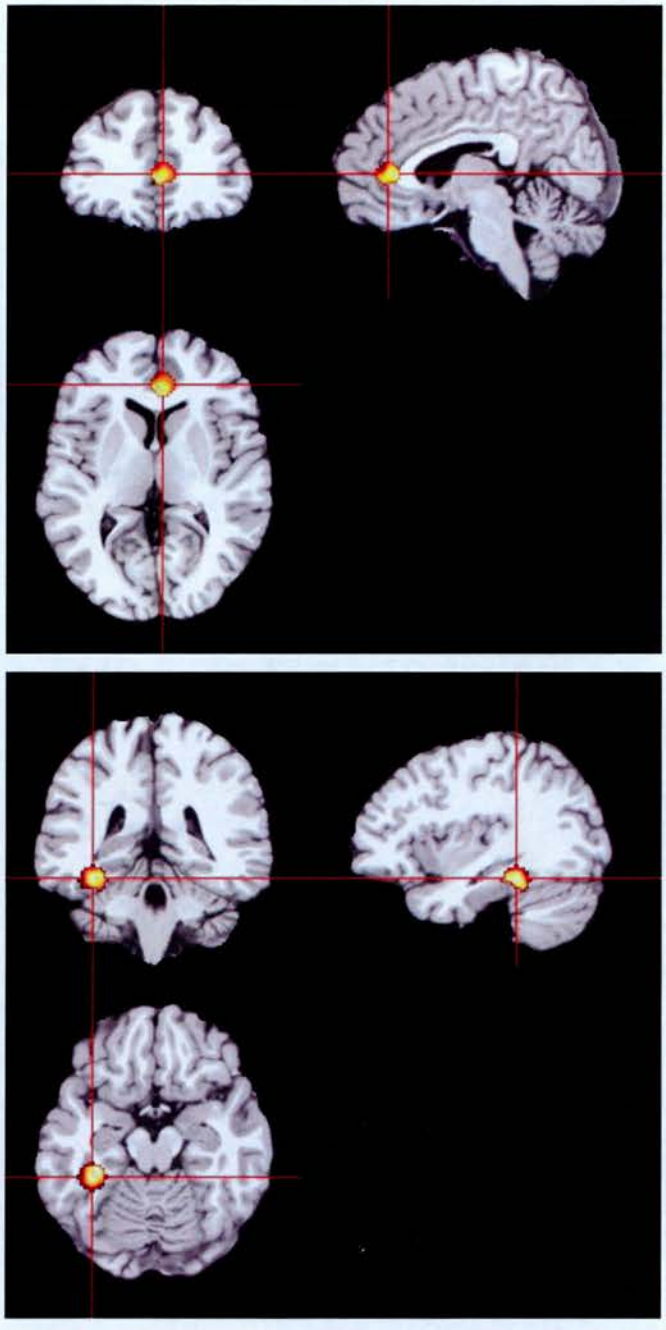


Figure 5.6 Random effects analysis showing activation within the rostral anterior cingulate and posterior parahippocampal gyrus predefined regions of interest which correlate positively with Hamilton score for the patient group, and additionally differ between patient and control groups. Regions (table 4) are only significant ($p<0.05$) after correction for multiple testing using a small volume correction.

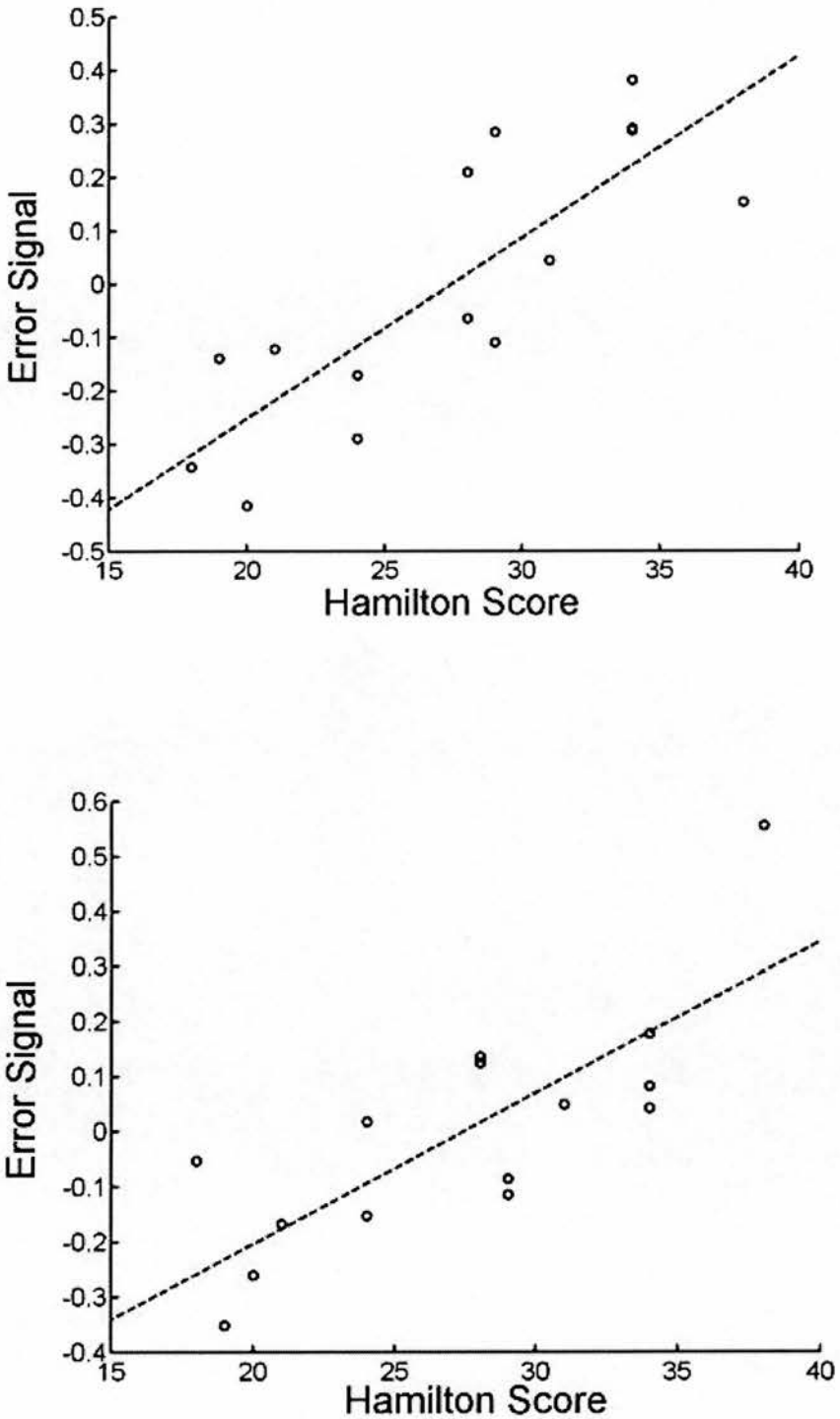


Figure 5.7 Patient group correlation of error signal with Hamilton score for rostral anterior cingulate (top) and left posterior parahippocampal gyrus (bottom) for regions shown in figure 5.6.

5.3.2 Region of Interest Connectivity Analyses

Effective Connectivity Analysis An independent t-test revealed various differences between the patient and control path values as shown in table 5. Considering the patient group alone, several paths correlated with Hamilton depression score, this also being listed in table 5. These changes in effective connectivity are summarised in figure 5.8. *Functional Connectivity Analysis* When the correlations between the hippocampus and other regions of interest were examined for both hemispheres, no significant differences were found.

5.4 DISCUSSION

5.4.1 Summary and Conclusions

The fixed effects analysis reported in this study has the advantage of higher statistical power than the random effects analysis, but only allows inference for the particular subjects taking part (Elliott, Friston & Dolan, 2000, p6161). It shows that a Kalman filter derived error signal was present in many brain regions of both patients and controls. The random effects analysis addressed the issue of comparison between the groups, allowing also generalisation to the patient and control populations. It showed that group differences in the predictive error signal existed in the rostral anterior cingulate and posterior parahippocampal gyrus regions. Additionally, the error signal correlated with illness severity for the patient group alone. The effective connectivity analysis indicated that the influence of one neuronal region on another also differed with respect to the error signal, between groups. In some cases, this difference additionally correlated with illness severity within the patient group.

The rostral anterior cingulate has often been reported as having abnormal function and structure in imaging studies of depressive illness (Ebmeier & Kronhaus, 2002). The region selected for investigation in this study corresponded to the most likely location for activation to be reported in diverse emotion induction studies of

Path			p
a	R	+G	0.021
b	R	-H	0.045
c	L	-G, +H	0.033, 0.041
d	L	+G	0.066
e	R	-H	0.006
h	L	+G	0.014

Table 5.5 Difference in effective connectivity; increased (+G) and decreased (-G) path value in patient group compared with controls, and positive (+H) and negative (-H) correlation with Hamilton depression rating for patient group alone. Path d almost reaches significance at a conventional $p<0.05$ level.

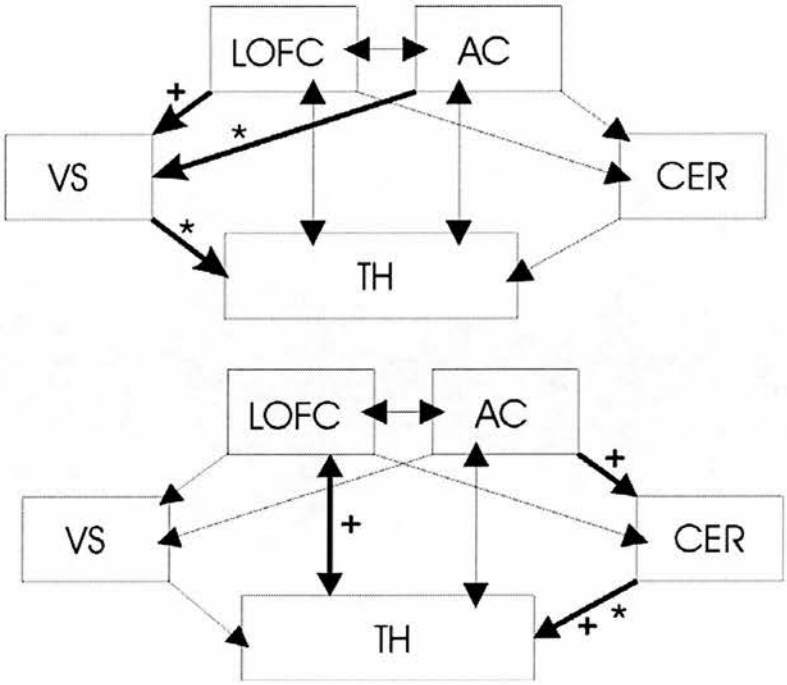


Figure 5.8 Difference in effective connectivity between patients and control groups (+) and correlation with Hamilton depression rating for control group only (*) for paths detailed in table 5.5. Right hemisphere (top) and left hemisphere (bottom) results are shown.

healthy subjects over the past decade (Steele & Lawrie, 2004b). This region is superior to, but continuous with, the subgenual anterior cingulate reported to have abnormal function and structure in a number of studies on depressive illness (e.g., Drevets, Price, Simpson, *et al*, 1997). Unfortunately, the susceptibility artefact prevented investigation of the subgenual region. However, reports of anterior cingulate abnormality in depressive illness are not confined to the subgenual region (Ebmeier & Kronhaus, 2002). Neural error signals reported for the anterior cingulate have usually been interpretated as reflecting various aspects of behavioural performance, which may involve prediction, but not necessarily motivational outcomes (Schultz & Dickinson, 2000); though see more recent work (Holroyd, Nieuwenhuis, Yeung, *et al*, 2003). This study provides evidence of the rostral anterior cingulate (emotion region) representing a predictive error signal which differs in depressive illness, and may be directly related to emotional experience and motivation. Other brain regions, such as the adjacent dorsolateral cortex, have been reported to exhibit error representation concerning both reward prediction and behavioural performance, though probably in different neuronal populations (Schultz & Dickinson, 2000). Both such representation might exist in the anterior cingulate, given increasing evidence of segregation of emotional and cognitive function (Bush, Luu & Posner, 2000; Steele & Lawrie, 2004b).

Considerable experimental evidence in animals has been accumulated, which is consistent with the septo-hippocampal system acting as a comparator based control system (Gray & McNaughton, 2000), predicting the next perceptual state, comparing this with the actual state, and generating an error signal when an event fails to occur or an unpredicted event occurs. Unlike other influential theories (O'Keefe & Nadel, 1978), Gray's theory is not confined to visuo-spatial prediction, specifically including predictions of pain, punishment or non-reward, and has a close relationship with anxiety. As discussed above, two previous studies have suggested that the hippocampus functions as a Kalman filter (Bosquet, Balakrishnon & Honavar, 1999; Dayan, Kakade & Montague, 2000). In studies of depressive illness, abnormal function and structure of the hippocampus has been reported on a number of occasions (Ebmeier & Kronhaus, 2002). Additionally, there is evidence that various forms of stress may adversely affect the structure of the hippocampus and it has been

suggested that antidepressants may have a neuroprotective effect (Reid & Stewart, 2001). Nevertheless, the mechanism by which such effects may be related to the symptoms of depressive illness are unknown. This study found evidence of a Kalman filter derived predictive error signal in the parahippocampal gyrus of control subjects which differed systematically from the patients. It suggests that the reported abnormalities of the hippocampal system in depressive illness may be linked to an abnormality of associative learning directly affecting emotional experience.

The link between Rolls' theory of emotion (which is not limited to omitted reinforcers) and predictive error signals was previously discussed. Rolls notes that his definition provides substantial opportunity for cognitive processing in emotions, such that cognitive processes will very often be required to determine whether an environmental stimulus is reinforcing (Rolls, 1999, p65). He notes that normally an emotion consists of cognitive processing, which results in a decoded signal that the event is reinforcing, together with the mood state. This conforms to a conventional distinction in descriptive psychopathology, that a mood is an emotion without an object (Sims, 1995). The other conventional distinction is that mood is prolonged whereas emotion (or affect) is not (Sims, 1995), and it is the sustained quality which is a defining feature of mood disorder (Sims, 1995). This is consistent with a definition of mood which combines the above: a mood state is a sustained "*predisposition*" (Owens, McKenna & Davenport, 2004, p230) to experience a given group of emotions for a given stimulus; i.e., mood reflects a *bias* with respect to the euthymic state. To the extent that emotions are directly related to reinforcers and error signals, it was this mood related bias that was investigated here.

This investigation of neural predictive error signals views brain function from the perspective of control theory in which a direct link between the error signal and emotion is postulated (chapter 4). The mechanisms of emotional regulation are increasingly recognised as relevant to the study of mood disorder (Gross, 1998). Mayberg has suggested that depressive illness involves failure of "homeostatic emotional control in times of increased cognitive or somatic stress" (Mayberg, 2003). McEwen has similarly discussed brain structural and endocrine abnormalities in mood disorder in terms of "allostatic load", where allostasis is defined as "maintaining stability or homeostasis through change" (McEwan, 2003). Such

contemporary theories of depressive illness, as reflecting failure if homeostatic mechanisms, echo much earlier concepts by Selye (Selye, 1950), and were anticipated by Solomon and colleagues a quarter of a century ago (Solomon, 1980a). Mayberg has further suggested that depressive illness consists of the following (Mayberg, 2001): degeneration of mesencephalic monoamine neurones and their cortical projections, remote changes in the basotemporal limbic regions, which could include the amygdala, anterograde or retrograde disruption of basal ganglia thalamocortical circuits, and secondary involvement of serotonergic neurones via disruption of orbitofrontal outflow. Chapter 6 describes an investigation into a possible midbrain structural abnormality which had been postulated to involve the monoamine neurones.

A limitation of this study is that there has been no attempt to take account of the detailed forms of the error signal. For example, dopaminergic neurones have distinct responses at different time points within a trial; e.g., at the time of stimulus presentation, point of action choice, and receipt of outcome (Schultz & Dickinson, 2000). The predictive error signal reported for other neuromodulators differs from dopamine. For example, noradrenergic neurones respond to unpredicted but not to predicted rewards (Schultz & Dickinson, 2000, p482). Additionally, differences between error signals indicate that anatomical structures process information in different ways; e.g., dopamine neurones emit a reward teaching signal without indicating the specific reward, striatal neurones adapt expectation activity to new reward situations, and orbitofrontal neurones process the specific nature of rewards (Schultz, Tremblay & Hollerman, 1998, p421). In this study, a single type of predictive error signal was calculated for the whole study period and entire brain. More detailed modelling of the error signal (e.g., O'Doherty, Dayan, Friston, *et al*, 2003) may allow clarification of specific differences between depressed patients and controls.

Whilst dopamine, noradrenaline and acetylcholine neuromodulator systems have been reported to exhibit typical predictive error signals, the serotoninergic system has not. Since selective serotonergic inhibitors (SSRIs) are effective treatments for many depressed patients, this might be considered evidence against the hypothesis, that depressive illness consists of dysfunction of the mechanisms of

associative learning. However, Daw and colleagues describe a theory of opponent interactions between dopamine and serotonin (Daw, Kakade & Dayan, 2002). They begin with a review of the experimental evidence for serotonin being associated with behavioural inhibition, aversive experience and punishment and follow it with a discussion of the experimental evidence supporting opponent interactions between dopamine and serotonin. They develop a quantitative theory of opponent interactions based on temporal difference theory. Importantly, they distinguish between short term phasic dopamine and serotonin signals, and long term tonic signals. Specifically, they suggest that a phasic serotonergic signal may report prediction error for future punishment and a tonic signal for reward. Dopamine phasic and tonic signals have a mirror opposite role.

This theory might begin to account for some otherwise unrelated observations; e.g., an acute increase in serotonin level being associated in animals with behavioural inhibition and punishment and in some patients with aversive anxiety "onset effects" when commencing antidepressants (BNF, 2004), yet when taken long term, the same antidepressants are effective treatments for depressive illness and a wide variety of anxiety disorders (BNF, 2004). Additionally, alcohol, stimulants and opiates are misused because people enjoy the acute effects, and it is well known that a common action of these otherwise diverse substances is to release dopamine in various brain regions (e.g., Steele & Lawrie, 2004a). In regular long term use all these substances are associated with an increased prevalence of depressive illness (McIntosh & Ritson, 2001) and substance misuse frequently complicates the treatment of depressive illness in general. The issue of whether patients become depressed as a consequence of excessive use of these substances, or attempt to medicate (perhaps pre-existing) symptoms of depression with the drugs, continues to be debated in the clinical literature. Daw and colleagues theory suggests that both may be correct. Short term use may alleviate some symptoms but repeated use may cause them (see quantitative modelling below). Of interest would be an experimental test of Daw and colleagues' model. Specifically, is there a mirror opposite effect of short and long term administration of an SSRI, and separately a stimulant, on the predictive error signals occurring during associative learning ?

This highlights another limitation of the current study. Most patients were receiving medication and none of the controls were. No relationship between amount of medication and illness severity was found. However, given the lack of an accepted method of estimating equivalent doses of different medications, it remains possible that such a relationship existed. Complicating recognition of this is the fact that subjects can vary widely in the extent of drug metabolism. Plasma monitoring of antidepressant levels was not done and would be necessary to accurately compare drug levels affecting the brain. Since there is no evidence to the contrary, the observed correlations of error signal with illness severity make it appear quite unlikely that the results of this study were due to a medication effect. However, such an effect can not be completely excluded. It is of course well known that different antidepressants have innumerable effects which are not captured by a simple relative dose magnitude index: e.g., tricyclics appear to exert their effects via a combination of serotonergic and noradrenergic reuptake inhibition plus variable receptor blockade, SSRIs are far more selective in the reuptake inhibition of serotonin, and MAOIs inhibit the breakdown of monoamines with relatively few effects on reuptake mechanisms, though also variable receptor blockade. Consequently, it is important to establish the effects of antidepressant medication on predictive error signals by a combination of studies on aminals and imaging studies on humans.

Ideally no patients should be receiving medication. This is because there are reports of structural and functional brain change apparently as a consequence of medication, and antidepressant medication is in fact expected to affect the signals of interest. It is very difficult to recruit patients who are not already receiving antidepressant medication. It is possible to acquire unmedicated patients but it can take a great deal of time; e.g., in two studies in which I have recruited a total of 30 unipolar depressed patients (Steele, Glabus, Shajahan, *et al*, 2000; Steele, Meyer & Ebmeier, 2004), this taking a total of over four years (in the context of other work), only two suitable unmedicated patients were identified. There are ways to speed up recruitment of such patients. For example, advertisements may be placed describing various symptoms and inviting people to contact the study organisers: respondents are then assessed to determine whether they satisfy diagnostic criteria for depressive illness. However, concern has been raised about how representative such

respondents are with regard to the population of patients of interest. Consequently this was not done, although the approach seems popular in North America.

The current analysis has investigated the error signals without taking account of context: i.e., whether the signal was particularly associated with a win or lose situation. The random effects analysis identified increased error signals in depressed patients compared with healthy controls, regardless of context. A further analysis is therefore planned, to determine whether the increased error signal in patients is particularly associated with winning or losing.

A further limitation of the current study is that there has been no investigation of the relationship between behaviour and brain activity. Such an investigation is limited with this paradigm since there is no meaningful performance by an individual subject (Elliott, Friston & Dolan, 2000). Nevertheless, it may still be possible to explore relative decision and latency issues between groups; e.g., patients might be more likely to switch response on trials following a loss, or may be insensitive to such a loss. Additionally, there may be a loss of normal "complexity" (chapter 4) of behavioural response and observed brain activity in patients. Therefore, we plan a future analysis of the data exploring these, and other potential links, between behaviour and error signal activity.

In summary, considerable neurobiological evidence supports the existence of predictive error signals in various brain regions reported to have abnormal function in many studies of depressive illness. This imaging study also found evidence of such a signal. Also, it was found to differ between patients and controls and vary systematically with depressive illness severity, supporting the initial hypothesis. The significance of these findings has been discussed in the context of relevant quantitative theories of brain function, clinical features of depressive illness and treatment.

5.4.2 Future work

This study only considered a small group of subjects and one game paradigm. Attempts at replication are therefore required, which might also address some of the limitations of the current study. There are several other areas which might also be explored in future work.

5.4.2.1 Delayed Antidepressant Response

As discussed above, Daw and colleagues have described a theory of opponent interactions between dopamine and serotonin, which may be relevant to mood disorder. It is therefore of interest to consider Solomon's related opponent process theory, with regard to quantitative modelling of the time course of mood disorder.

It is possible to model Solomon's description by calculating "a" and "b" process functions as convolutions of a square wave (stimulus) with a gaussian of given full width half maximum and phase delay (c.f., information theory, Shannon, 1948), together with an exponential decay term for the "b" process. The result is shown in figure 5.9 which appears qualitatively indistinguishable from the original figure 4.5. Solomon and colleagues discussed their theory with regard to substance misuse, which involves repeatedly administering a substance before the "b" process has decayed (chapter 2). The effect of this type of dosing is shown in figure 5.10. If the "b" process area is initially less than the "a" process area, then mood is initially elevated but soon falls below euthymic baseline, with strengthening of the "b" process. On discontinuation of substance misuse, mood gradually recovers.

Now consider the case of a patient with a depressive illness, not misusing substances, who begins to take an antidepressant such as an SSRI: figure 5.10. Based on evidence that an *acute* increase in serotonergic activity is associated with increased anxiety and aversive experiences (chapter 2), this can be modelled by a square wave pointing in the *opposite* direction to that for a stimulant. If the "b" process area is less than the "a" process area, there is an initial worsening of mood (c.f., "onset effects" or the initial stages of "desensitisation") which subsequently disappears, and mood recovers to a euthymic state. Figure 5.10 assumes that the "b" process does not continue to strengthen once a euthymic state has been reached: susceptibility to this situation may distinguish bipolar from unipolar mood disorder. If the SSRI is discontinued before the underlying illness has resolved, mood falls again to the original depressed state (figure 5.10).

Only a very brief discussion has been possible here to indicate the direction of possible future work. This could consist of fitting the model to clinical data, such as the time course of recovery from a depressive illness, which may allow

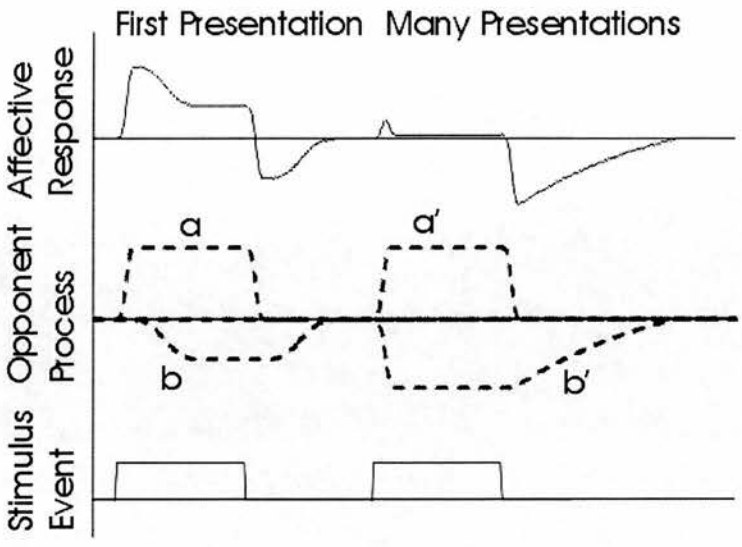


Figure 5.9 Solomon's opponent process affective response. All parts of this diagram have been generated using the mathematical model described in the text (calculations done in Matlab). The original curves described by Solomon and colleagues appear essentially indistinguishable (chapter 2).

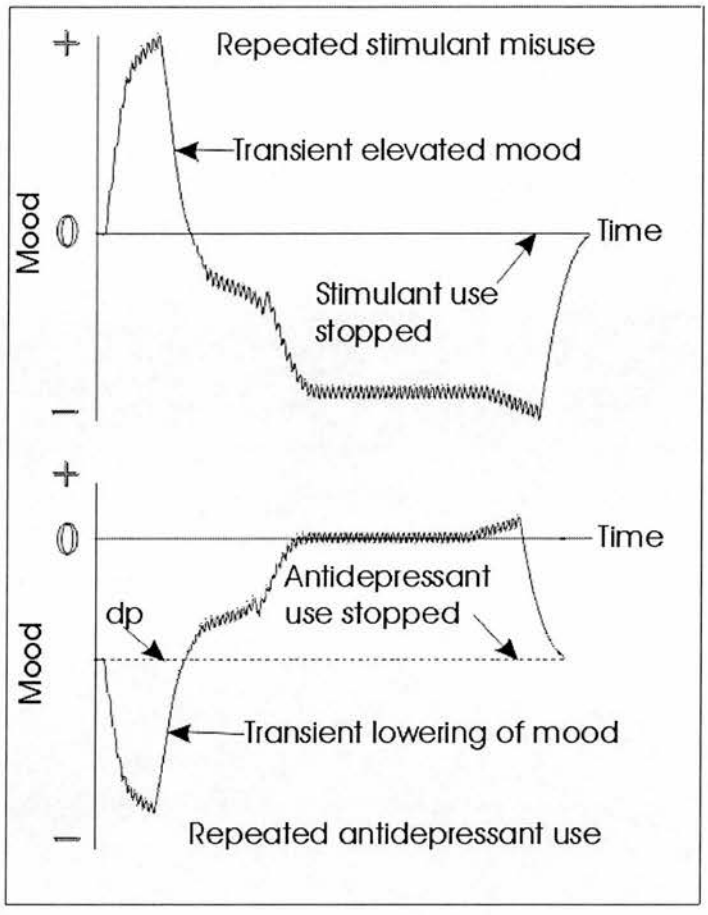


Figure 5.10 The effects of repeat dosing of a substance before the "b" process has decayed. Mood (traditionally defined as a predisposition to emotion along an elation (+) to sadness (-) dimension, Owens, McKenna & Davenport, 2004, p230) is represented by the vertical axis, time by the horizontal axis. Calculations have assumed a simple linear progression from the "first presentation" to "many presentations" state, the latter being then kept constant until stopping the substance use. The top diagram shows the effects of repeated use of a substance such as a stimulant. The lower diagram shows the opposite effect, for the case of a non-substance misusing depressed individual repeatedly using an antidepressant. Alterations in receptor expression (e.g., Bonhomme & Esposito, 1998; Harrison-Read, Tyrer & Sharpe, 2004, p460) may in part underlie the "b" process change over time. Abbreviations: depression severity (dp), euthymic state (0). According to this model, use of a stimulant could transiently improve mood during a depressive illness at the price of worsening mood in the long run.

predictions about underlying temporal difference error signals. This might be tested with a suitable fMRI study.

5.4.2.2 Large Error Signals and Physical Treatments

The study of depressed patients discussed above found evidence for a significantly increased predictive error signal. In the treatment of Parkinson's disease, it is generally accepted that limited lesions of small parts of the basal ganglia, or direct electrical stimulation of the *same* structures can alleviate symptoms(Baev, Greene, Marciano, *et al*, 2002). Two questions are: how can destroying part of a network (motor BGTCL) improve function, and even more puzzling, how can electrical stimulation of the same structure, achieve the same result as a lesion ? Non-stochastic theories of brain function have difficulty accounting for these observations.

Baev and colleagues have considered these questions in terms of stochastic control theory (Baev, Greene, Marciano, *et al*, 2002). They argue that in the case of disease, the model prediction does not match optimally with reality, which generates a *large error signal*. Partial lesioning of such a model reduces the precision of the prediction of object states (with the gaussian of predictions in "state space" becoming lower and wider) consequent increased overlap with reality, and so a diminished error signal (Baev, Greene, Marciano, *et al*, 2002). In the case of deep brain stimulation, the electrical stimulation does not convey information; i.e., it is noise. This makes it easier for a network to achieve a global optimal condition, and not get stuck in a less optimal "local minimum" (Baev, Greene, Marciano, *et al*, 2002) (c.f., simulated annealing). Baev and colleagues argue that stimulation may consequently result in a better match with reality, and thus a reduced error signal, but also at the expense of reduced predictive precision (Baev, Greene, Marciano, *et al*, 2002).

This suggests a mode of action for ECT (and repetitive TMS, plus vagal nerve stimulation). Even more controversially, it also suggests a mode of action for neurosurgery for mental disorder (Freeman, Crossley & Eccelston, 2000), which also involves selective lesions of a BGTCL (limbic rather than motor): figure 5.11. Indeed, therapeutic effects of both lesioning (CRAG Working Group, 1996) and

stimulation (Nuttin, Gybels & Meyerson, 1999) of the *same* structure (anterior internal capsule, figure 2.1; through which the limbic reciprocal corticothalamic projections pass) have been reported. It also implies a circumstance when stimulation will not work (no global minimum in the brain network corresponding to a euthymic state), and a circumstance when lesions will not work (predicted states too far from reality) . These issues could be explored with network simulations and imaging studies.

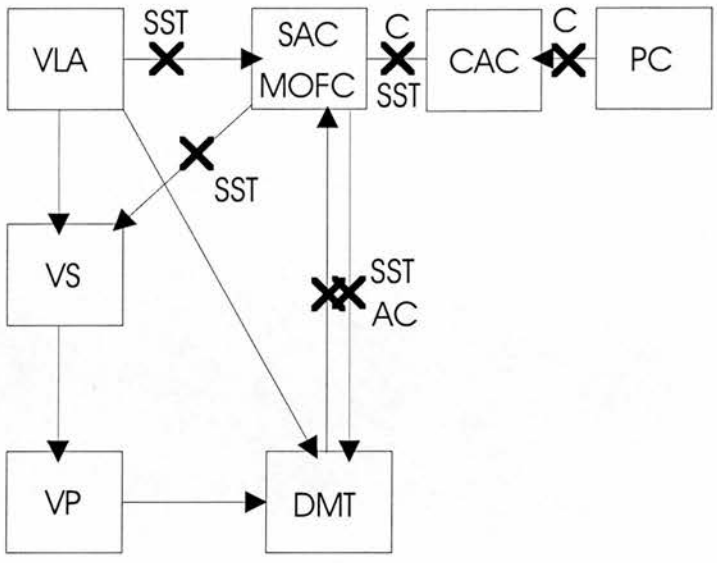


Figure 5.11 Limbic BGTCL with superimposed neurosurgery for mental disorder (NMD) lesion locations. Abbreviations: PC posterior cingulate, SAC subgenual anterior cingulate (see figure 2.5), MOFC medial orbitofrontal cortex (see figure 2.4), CAC caudal anterior cingulate, VLA ventrolateral nucleus of the amygdala, VS ventral striatum, VP ventral pallidum, DMT dorsomedial thalamic nucleus, SST stereotactic subcaudate tractotomy, C cingulotomy, AC anterior capsulotomy. Limbic loop circuitry based on various reviews (Alexander & Crutcher, 1990; Price, 1999). Lesion locations derived from various descriptions (Ballantyne, Bouckoms, Thomas, *et al*, 1987; Knight, 1964; Meyerson & Mindus, 1988; Richardson, 1973). The cingulotomy lesion location is variable and often deep to caudal anterior cingulate (see figure 2.5).

CHAPTER 6

INVESTIGATION INTO A POSSIBLE STRUCTURAL ABNORMALITY OF THE BRAINSTEM OF PATIENTS WITH UNIPOLAR DEPRESSION

6.1 INTRODUCTION

Virtually all empirically derived antidepressants are known to increase monoamine levels. This observation led to the well known monoamine hypothesis of mood disorder (Bloom & Kupfer, 1995; Schildkraut, 1965). The nuclei of cells synthesising serotonin, noradrenaline and dopamine are located in the brainstem, and axonal projections from these nuclei to other brain areas form white matter tracts such as the medial forebrain bundle (chapter 2). The anterior cingulate has often been reported as having abnormal function and structure in depressive illness (Ebmeier & Kronhaus, 2002). Studies on primates indicate that there is a topographical projection from the cingulate to the pons. The anterior cingulate projects to the midline pons and the posterior cingulate to the lateral pons (Vilensky & van Hoesen, 1981). Although virtually all regions of the anterior cingulate have been reported as abnormal in depressive illness, the subgenual region has been a particular focus of reports (Drevets, Price, Simpson, *et al*, 1997). In primates, the subgenual anterior cingulate projections to the brainstem are restricted to the posterior midline periaqueductal grey matter and raphe serotonergic nuclei (Freedman, Insel & Smith, 2000).

A series of studies using transcranial ultrasound have reported a structural abnormality of the midbrain midline of patients with unipolar depressive illness compared with controls (Becker, Becker, Struck, *et al*, 1995; Becker, Struck, Bogdahn, *et al*, 1994b). The same structural abnormality has also been reported when depressed patients have been compared with non-depressed patients, both having a variety of neurological diseases; e.g., Parkinson's disease (Becker, Becker, Seufert, *et al*, 1997; Berg, Supprian, Hoffmann, *et al*, 1999), dystonic syndromes (Naumann, Becker, Toyka, *et al*, 1996) and Huntington's disease (Becker, Berg, Lesch, *et al*, 2001) but not multiple sclerosis (Berg, Supprian, Thomae, *et al*, 2000). The structural abnormality was reported to occur in unipolar depressed patients, was

unrelated to severity of current illness, and was absent in patients with bipolar disorder and schizophrenia (Becker, Becker, Struck, *et al*, 1995). In all cases the abnormality was reduced echogenicity of the midline.

These ultrasound investigations have been supplemented by T_2-weighted MRI studies. *Increased* intensity of the midline has been reported for unipolar depressed patients when compared with bipolar patients and controls in a retrospective study using T_2-weighted MRI (Becker, Becker, Berg, *et al*, 1998). Additionally, a prospective study of depressed patients with Parkinson's disease reported a shift of signal in the midbrain using a semi-quantitative assessment of relaxation time (Berg, Supprian, Hoffmann, *et al*, 1999). It has been suggested that reduced echogenicity of the midline and increased T_2-weighted signal intensity may be consistent with disruption of myelinated fibre tracts running along the brainstem, which include the monoaminergic tracts (Becker, Berg, Lesch, *et al*, 2001). There are few post-mortem studies of the brainstem of patients who suffered from major depression. Nevertheless, one small study has reported that patients with unipolar depression have a distinct disruption of the mesencephalic fibre tracts which may be consistent with both the ultrasound and MRI findings (Becker, Berg, Lesch, *et al*, 2001). It has been suggested that such an abnormality could result in a reduction of brain monoamines and be reversed by antidepressants (Becker, Berg, Lesch, *et al*, 2001). Clearly then, the collective replicated ultrasound, MRI and histopathological studies are an important work with major implications for understanding the potential causes of unipolar depressive illness.

A literature search did not reveal any reports of attempted replication by independent groups. Therefore, the aim of this study was to attempt replication of these findings using transcranial ultrasound imaging and explore the use of a new technique, DT-MRI. The latter may be the imaging modality of choice, since it obtains information on the structural integrity of white matter tracts which can be analysed in an objective automated manner. In contrast to ultrasound imaging, it does not rely on the clinical skill of an operator trying to record a subtle abnormality in a low signal/noise image, and removes the possibility of knowledge of the patient's condition introducing bias. Although ultrasound images may be analysed independently and blindly after acquisition, the acquisition process itself (and hence

images) could be influenced by the operator being aware of the patient's appearance and guessing their clinical condition.

A power analysis was done for the planned ultrasound investigation. For the first pilot study, mean echogenicity scores of 1.3 (0.47) and 2.8 (0.64) for the patient and control groups respectively were reported (Becker, Struck, Bogdahn, *et al*, 1994b), where the number in brackets is the standard deviation. For the later study, echogenicity scores of 1.4 (0.6) and 2.8 (0.5) were reported (Becker, Becker, Struck, *et al*, 1995). These correspond to Cohen's d effect sizes (Cohen, 1988) of 2.68 and 2.54 respectively, which are extremely large; by convention, 0.2 is regarded as small, 0.5 medium and 0.8 large. Based on an effect size of 2.5, alpha of 0.05, 15 patients and 15 controls, the power of this current study was calculated using the Gpower program (Erdfelder, Faul & Buchner, 1996) to be 100%.

To clarify the power value of 100% quoted above, it is of course not possible for a power to be precisely 100%. The Gpower program calculates the power to four significant decimal places and, e.g., for an effect size of 2.2, a power of 0.9999 is determined. In the case of the even larger effect size of 2.5, a power of 100% is calculated due to "rounding", which is a convention to avoid long lists of "9"s after the decimal point.

6.2 METHODS

6.2.1 Subjects

Permission for the study was obtained from the local ethics committee and written informed consent obtained from each subject. A total of 30 right-handed subjects participated in the study. One subject was unable to tolerate scanning resulting in data from 15 patients and 14 controls (table 5.1). The power of the ultrasound investigation was recalculated for 29 subjects and still found to be 100%. The same subjects also took part in an fMRI study described in chapter 5.

Subjects were excluded if they had a history of significant head injury or structural brain abnormality (including vascular disease), substance (including alcohol) misuse, physical disorder known in some cases to be associated with abnormal brain structure (e.g. epilepsy) or receiving medication which might alter brain structure (e.g. steroids). As a prerequisite for MRI scanning, subjects were

additionally excluded if they had deliberately (e.g. cardiac pacemaker) or accidentally implanted metal. Subjects had to be without serious physical disease. The upper age limit was not restricted to 65 years. All patients had an unequivocal diagnosis of recurrent unipolar major depressive illness. Patients were excluded if the diagnosis of the current or any previous illness was in doubt (e.g. a possible previous manic episode). Most patients were receiving a variety of medications (table 5.2) at doses similar to that described in a previous study (Becker, Struck, Bogdahn, *et al*, 1994b). Medication had remained at a constant level for at least two weeks prior to scanning. Patients were recruited from inpatients and outpatients at the Royal Edinburgh Hospital. Controls were acquired from work colleagues, friends and relatives. Potential controls were excluded if they admitted to a history of any previous psychiatric illness.

Details of the subjects are given in table 5.1. The control group was matched on the basis of age, NART (Nelson & Wilson, 1991) and percentage of male subjects. All subjects completed a BDI (Beck, Ward & Mendelson, 1961) as a screening test. Patients satisfied DSM IV criteria for major depressive disorder and a Hamilton depression rating (Hamilton, 1960) was obtained as an index of depression severity. The ratings and scans were obtained on the same day and time of day (late morning) because of diurnal variation in mood which was present in some patients.

6.2.2 Ultrasound Image Acquisition and Analysis

Ultrasound images were obtained from subjects using a phased array system equipped with a 2 MHz transducer (Siemens Elegra, Germany). All images were acquired by the same author (JMW), a consultant neuroradiologist with extensive experience of transcranial, including brainstem ultrasound, scanning. Images were obtained blind to subject details. Additionally, subjects were told not to speak or otherwise communicate with JMW before and during scanning. Depressed patients typically look unwell in their general demeanour, and so it is difficult to achieve full and effective blinding. This was considered relevant, since moving the ultrasound probe by a small amount can cause a midline echo to disappear, and the probe is unstable since it is hand-held. Therefore, as a check on the effectiveness of blinding,

JMW was asked upon completion of each scan session to guess whether the subject was a patient or control.

Images of the midbrain with red nucleus and rostral pontine brainstem were obtained in an axial scanning plane through a preauricular acoustic bone window (Becker & Griewing, 1998; Becker, Struck, Bogdahn, *et al*, 1994a). Acquisition was standardised such that the contralateral skull surface was just visible resulting in a typical penetration depth of 14 cm. The dynamic range was 28 dB with consequent high tissue contrast. Typically, around 7 images were recorded from each subject which always included views obtained from both head sides. The objective was to record the clearest images of the midline echo from the lower midbrain. Upper midbrain images were not recorded since the third ventricle appears as a dual midline echo and can be confused with the single midline echo of interest (Becker, Becker, Berg, *et al*, 1998). Similarly, upper pontine images were not obtained since the semi-quantitative rating procedure described by Becker and colleagues requires direct comparison of the midline to red nucleus echos (Becker, Struck, Bogdahn, *et al*, 1994b; Becker, Mintun, Diehl, *et al*, 1994) - the red nucleus is not present in the pons. Digitised images were stored for later analysis in the standard Siemens Elegra format which conforms to a "tif" file with 888*666 pixels, 24.79*18.59 cm image size, 8 bits per pixel and 91*91 dpi resolution.

For analysis, the anonymised stored images were rated (by JMW) according to the semi-quantitative method described by Becker and colleagues (Becker, Struck, Bogdahn, *et al*, 1994b; Becker, Mintun, Diehl, *et al*, 1994). This is a four point scale of echogenicity of the brainstem raphe compared with the red nucleus: 1 – raphe not visible/isoechogenic compared with adjacent brain parenchyma, 2 – decreased echogenicity of the raphe as compared with the echogenicity of the red nucleus, 3 – normal/identical echogenicity compared with the red nucleus, and 4 – increased echogenicity and /or widened raphe structure. As with previous studies, the null hypothesis of no difference of midline echogenicity in patients compared to controls was investigated with a U test.

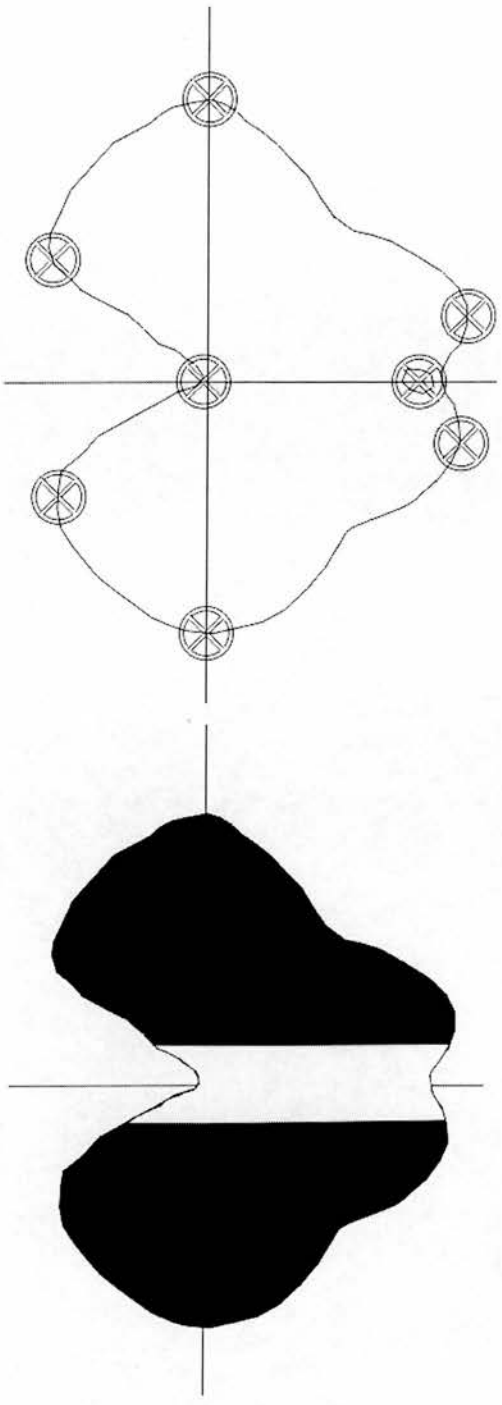


Figure 6.1 Lower midbrain template image shown at the top. Anterior is left. Circles indicate landmarks used for coregistration. The black region was used for normalisation of echo intensity.

Additionally though, to explore the possibility of localised change in echo intensity between the patient and control groups, for each subject a single scan was identified at the time of rating which showed the brainstem midline most clearly. Quantitative image analysis was done on these scans. A template image of the brainstem (figure 6.1) was constructed using a transverse slice through the lower midbrain of a high resolution T_1 weighted MRI scan (http://www.mrccbu.cam.ac.uk/Imaging/mnispace.html) which conforms to Montreal Neurological Institute (MNI) space. Landmarks were defined on the template which could also be identified on the ultrasound scans. Coregistration of the ultrasound images to the template was done (blind to subject details) using custom Matlab scripts, which made use of the Matlab image processing toolbox (The Mathworks, Natick, MA, USA). The latter includes routines for landmark based coregistration.

An affine transformation was used to coregister the images. The echo intensity of each image was normalised to the average intensity of the template image excluding a wide midline region (figure 6.1). The normalisation region included, but was not limited to, the red nucleus. A one group, two tailed t-test was used to test the null hypothesis of no difference in echo intensity at each pixel compared with the average intensity of the normalisation region. Correction for multiple testing was done using the false discovery rate (FDR) method (Genovese, Lazar & Nichols, 2002) with a conventional threshold of $p<0.05$. In a similar manner, a two group independent t-test was calculated to test the two tailed null hypothesis of no difference in average midline echo intensity of patients compared with controls.

6.2.3 DT-MR Image Acquisition and Analysis

All MR imaging data were obtained using a GE Signa LX 1.5 T (General Electric, Milwaukee, WI, USA) research-dedicated scanner, equipped with a self-shielding gradient set (22 mT/m maximum gradient strength and 120 T/m/s slew rate) and manufacturer-supplied 'birdcage' quadrature head coil. Each subject underwent an axial T_2-weighted fast spin-echo (FSE) sequence to identify silent brain pathology and a T_1-weighted volume scan. This was followed by a DT-MRI

protocol specifically designed to image the brainstem. The duration of this examination was approximately 15 minutes.

For DT-MRI acquisition, diffusion-weighted (DW) images were acquired from 11 slice locations covering the region from the pons to the body of the lateral ventricles using a single-shot spin-echo echo-planar (EP) imaging sequence. The center slice was aligned with the midbrain-pontine junction. Sets of axial DW-EP images ($b = 0$ and 1000 s/mm^2) were collected with diffusion gradients applied sequentially along six non-collinear directions (Basser & Pierpaoli, 1998). Seven acquisitions consisting of a baseline T_2-weighted EP image and six DW-EP images, a total of 49 images, were collected per slice position. The acquisition parameters for the DW-EP imaging sequence were 11 contiguous axial slices of 4 mm thickness, a field-of-view of 180 × 180 mm, an acquisition matrix of 96 × 96 (zero filled to 256 × 256), a TR of 8.0 s and a TE of 98.6 ms. All the DICOM format magnitude images collected in each examination were transferred from the scanner to a Sun Blade 2000 workstation (Sun Microsystems, Mountain View, CA, USA) and converted into Analyze (Mayo Foundation, Rochester, MN, USA) format using "in house" software written in C. The following computations were then performed using the Matlab programming environment.

The set of seven component DW-EP images for each gradient direction were averaged to give seven high signal-to-noise ratio images for each slice location. Geometric image distortions arising from the strong eddy currents created by the diffusion gradients were then corrected in the six averaged DW-EP images using a modified version of the iterative cross-correlation algorithm (Bastin & Armitage, 2000). Within each voxel the six elements of the apparent diffusion tensor of water (**D**) and the T_2-weighted signal intensity were estimated by multivariate linear regression from the signal intensities measured in the DW-EP images (Basser, Mattiello & LeBihan, 1994). Diagonalization of **D** yielded the magnitude sorted eigenvalue (λ_i), maps of the T_2-weighted signal intensity and the fractional anisotropy (FA) (Basser, 1995)

$$FA = \sqrt{\frac{3}{2}}\sqrt{\frac{\left(\lambda_1 - < D >\right)^2 + \left(\lambda_2 - < D >\right)^2 + \left(\lambda_3 - < D >\right)^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \qquad 6.1$$

which were generated on a voxel-by-voxel basis and the resultant images converted into Analyze format. The FA measures the fraction of the total magnitude of **D** that is anisotropic, and takes a value of 0 for isotropic diffusion ($\lambda_1 = \lambda_2 = \lambda_3$) and 1 for completely anisotropic diffusion ($\lambda_1 > 0$; $\lambda_2 = \lambda_3 = 0$).

Statistical pre-processing and analysis of the DT-MR images were done using SPM99 (http://www.fil.ion.bpmf.ac.uk/spm/). For pre-processing, T_2-weighted EP images were spatially normalised to the SPM template using an affine transformation. Additional non-linear transformation was not done due to the limited number of slices. As a check on the spatial normalisation procedure, coordinates corresponding to anatomical landmarks in each T_2-weighted EP image were compared with the corresponding coordinates in the high resolution T_1 weighted image. No problems with normalisation were found. The T_2-weighted EP image normalisation parameters were then applied to the FA images, which were smoothed with an 8 mm isotropic gaussian filter. The null hypothesis of no difference between patient and control groups was finally tested with an independent two group t-test.

6.3 RESULTS

6.3.1 Ultrasound Imaging

A chi-square test indicated that the correct diagnosis of the subject was not guessed at more than by chance alone, p=0.53, (table 6.1). Table 6.2 shows the results of rating the anonymized recorded scans. The mean echogenicity scores were calculated as 2.07 (1.39) and 2.57 (1.16) for the patient and control groups respectively. This difference is not significant (U test, p=0.37). The score for the control group is very similar to Becker and colleagues' previous reports: 2.80 (0.5 and 0.64). In contrast, the patient score is different from previous reports: 1.3 and 1.4 (0.47 and 0.60). Cohen's d is 0.39, which is of small to medium size, and in the direction of previous studies. For an effect size of 0.39, an *a priori* power analysis was calculated using Gpower (Erdfelder, Faul & Buchner, 1996), assuming an asymtotic relative efficiency of 0.955 for a U test (Erdfelder, Faul & Buchner, 1996). For an alpha of 0.05, a study would have a power of 70% if it included a total of 173 subjects (2 tailed hypothesis) or 132 subjects (1 tailed hypothesis).

Figure 6.2 shows the average spatially coregistered and echo intensity normalised images for the patient and control groups. The hyperechogenic basal cisterns (BC) and aqueductal region (AQ) are visible as are hyperechogenic red nuclei (RN). The midline (M) structure also appears to be present in both the average images. Figure 6.3 shows the result of testing the 2 tailed null hypothesis of no difference between the average echo intensity of the normalisation region (figure 6.1) and any voxel in the image. The BC and AQ regions are significantly increased in intensity as are the RN on the sides of the brainstem closest to the probe. There are two small regions of significantly increased intensity in the midline (M) of the control group but not patient group, but far larger are the regions of significantly decreased (D) echo intensity lateral to the midline (figure 6.3). Consistent with the results of the semi-quantitative analysis, when the patient and control groups were directly compared using an independent 2 group t-test, no significant differences were found.

6.3.2 DT-MR Imaging

Table 6.3 shows the results of the comparison of patient and control groups. No region of significant change was found in the midbrain. Two regions of decreased FA (patients compared to controls) were identified outwith the brainstem. The first is located in the right lateral temporal lobe and may lie within a region of reduced grey matter reported for patients with treatment resistant depressive illness (Shah, Ebmeier, Glabus, *et al*, 1998). Inspection of the FA scans from each subject did not identify an obvious artefact in this region. The other significant region is located within the left uncinate fasciculus and might be related to previous reports of structural and functional abnormality of the left prefrontal and temporal lobes in depressive illness (Ebmeier & Kronhaus, 2002). However, this region does not remain significant after correction for multiple testing using the FDR method. No regions of increased FA were identified.

	Guess			
	Patient	Control	DK	Total
Patient	6	3	6	15
Control	3	3	8	14

Table 6.1 Guessed diagnosis of scanned subject (don't know; DK). The correct diagnosis of the subjects were not guessed at more than by chance alone.

	Score 1	Score 2	Score 3	Score 4	Mean ± SD
Controls	4	1	6	3	2.57 (1.16)
Patients	9	0	2	4	2.07 (1.39)

Table 6.2 Echogenicity of the brainstem raphe. The difference between patient and control groups is not significant.

	MNI coordinate
Right lateral temporal cortex	(64,-34,-26)*
Left uncinate fasciulus	(-24,10,-14)

Table 6.3 Regions of significantly reduced ($p<0.001$, uncorrected) FA of patient group compared with controls. (* indicates $p<0.05$ after correction for multiple testing using FDR method). No regions of increased FA were identified.

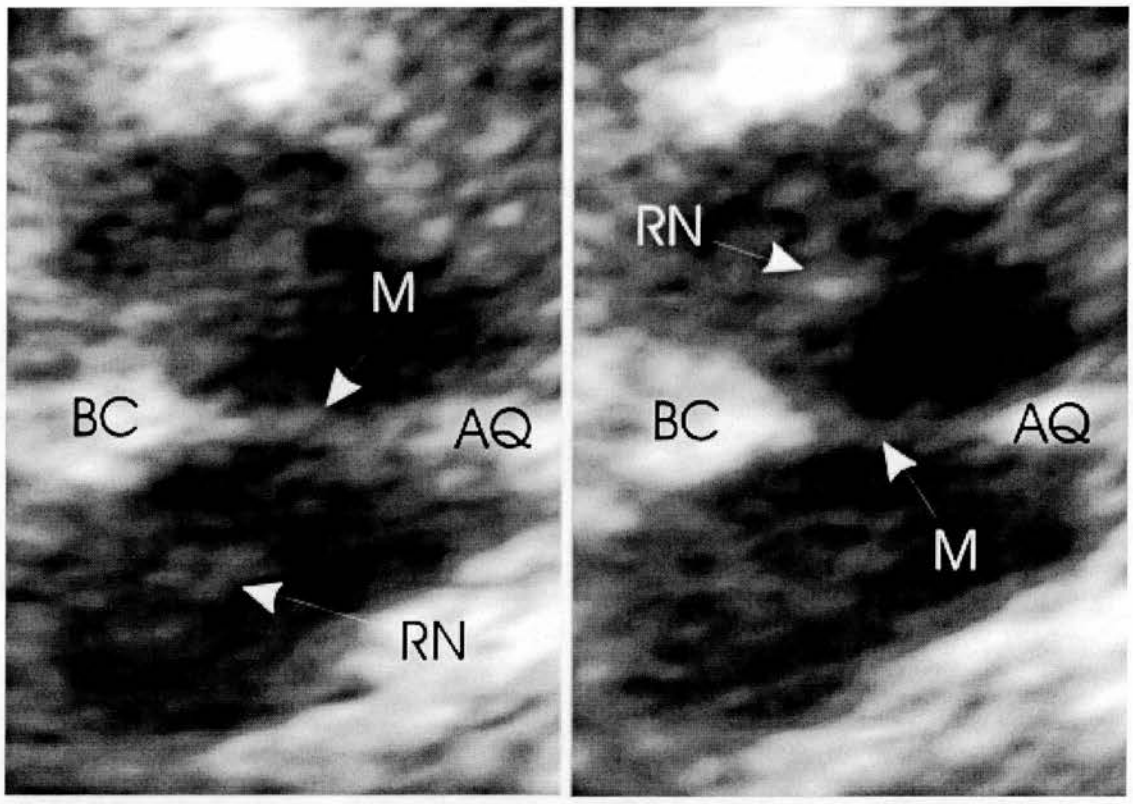


Figure 6.2 Average echo intensity normalised images for control group (left) and patient group (right). Anterior is left and probe scanned from top. Basal cistern (BC), aqueduct (AQ), midline of midbrain (M) and red nucleus (RN) are indicated.

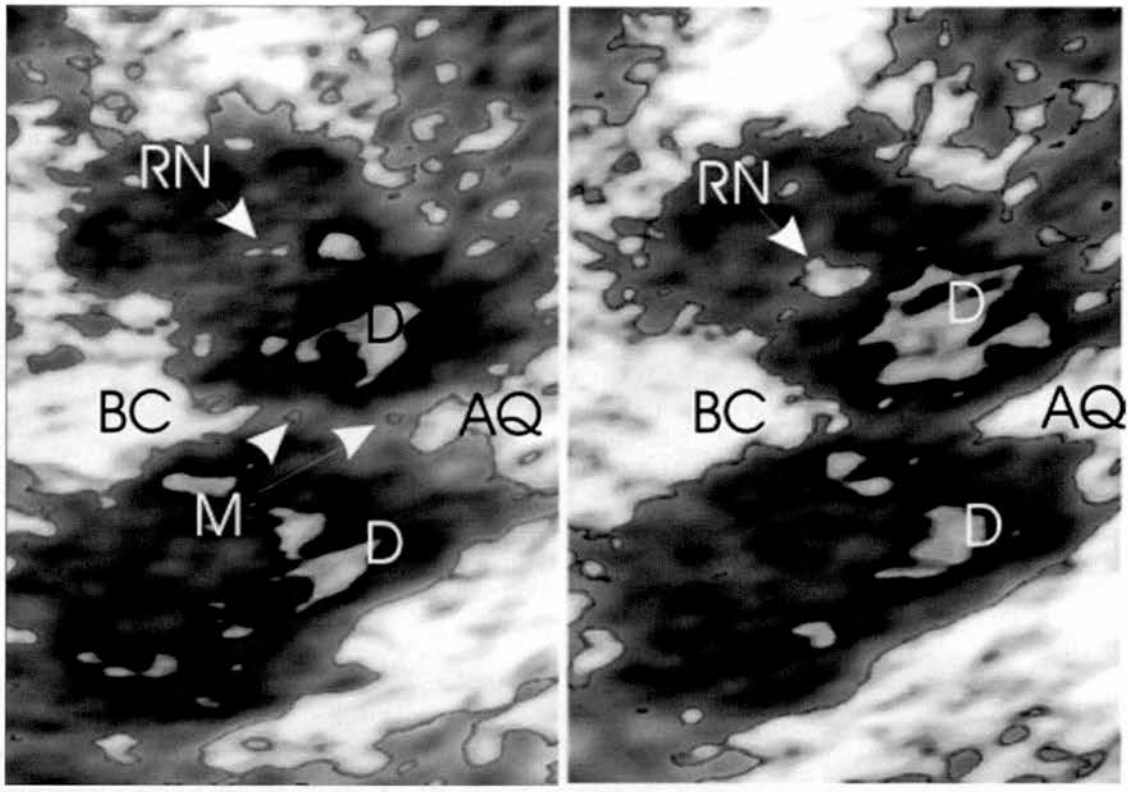


Figure 6.3 Average echo intensity normalised images for control group (left) and patient group (right). Anterior is left and probe scanned from top. Regions of significantly ($p<0.05$, FDR correction for multiple testing) increased and decreased echo intensity, compared to the mean of the normalisation region, are shown as a coloured overlay. Basal cistern (BC), aqueduct (AQ), midline of midbrain (M) and regions of significantly decreased echo intensity (D) are indicated.

6.4 DISCUSSION

6.4.1 Summary and Conclusions

As noted earlier, previous studies by Becker and colleagues reported extremely large effect sizes. Based on this, including 29 subjects had virtual certainty of rejecting the null hypothesis at $p<0.05$. However, no significant difference between patient and control groups was found. With regard to blinding, depressed patients typically look unwell in their general demeanour, and so it is difficult to achieve full and effective blinding. It is unclear how effective the blinding was in previous studies, since a check on blinding was not described.

The analysis of the coregistered ultrasound images from each group identified significant features within the brainstem, including statistically significant red nuclei echos. Consequently, it is unlikely that the images were of insufficient quality. This interpretation is additionally supported by the semi-quantitative analysis. Becker and colleagues' studies reported an echogenicity score for the control group similar to that which was found here. The main difference between this and previous studies, is that the patient score was not found to be approximately half the control score. An echogenic midline was found for a much higher percentage of patients than in the studies of Becker and colleagues. Inability to replicate previous results might be due to a difference between patient samples. However, the average Hamilton ratings of illness severity in previous studies were 22.8 (6.6) (Becker, Struck, Bogdahn, *et al*, 1994a) and 17.5 (8.4) (Becker, Becker, Struck, *et al*, 1995) which appears less than 27.3 (10.7) in this study. Nevertheless, Becker and colleagues did not find any correlation between illness severity and echogenicity of the brainstem. Consequently, it is unlikely that imaging more severely unwell patients explains the discrepancy.

If not due to illness severity, it might be due to comorbidity. Patients with comorbid substance (e.g., alcohol) misuse were excluded from our study. Alcohol misuse is very common in depressed patients (McIntosh & Ritson, 2001), and is generally recognised to be a cause of widespread structural brain changes (Kril & Halliday, 1999). It is unclear if similar patients were excluded in the previous

studies. In this study, consistent with the results of the ultrasound investigation, no difference in fractional anisotropy of the brainstem of patients was found.

In summary, we were not able to replicate Becker and colleagues' reports of a significant reduction in the echogenicity of the brainstem midline in unipolar depressed patients. This means that the abnormal predictive error signals identified in the same patients could not be accounted for by such a structural abnormality. Nevertheless, the ultrasound investigation indicated that there may be a trend in the direction of previous studies, suggesting that the magnitude of the effect is less than previously suggested, at least for the subject inclusion criteria adopted for this study. Given the implications of Becker and colleagues' reports with regard to identifying the causes of depressive illness, it is important that other groups attempt similar studies. The power estimates detailed earlier may be useful in this regard.

6.4.2 Future Work

Methods to quantitatively analyse DTI data are only now being developed. Whilst the approach used above has been used in other studies, it may not have had the highest statistical power, in this particular study. This was due to the need to balance the requirement for maximising spatial resolution (the brainstem is a small structure), against minimising noise level. However, a different method of image processing might help to reduce the noise level in the images, and so increase the power of the tests. Specifically, algorithms exist which use the information from each plane of the image to construct continuous fibre tracts (eg, Terajima & Nakada, 2002). It is possible that such fibre tract tracing might reduce the noise level in the image, and so potentially lead to higher power statistical analyses. Future work to investigate this issue is planned.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

7.1 SUMMARY AND CONCLUSIONS

Brain regions which are repeatedly reported as having abnormal function in depressive illness, such as the orbitofrontal and anterior cingulate cortices, amygdala and hippocampus, and the subcortical components of the ventral BGTCLs, are the substrate in animals for associative learning and the representation of the rewarding or aversive aspects of stimuli. Imaging studies on humans appear consistent with these studies. Normal human emotion has been linked to the presence or omission of reinforcers. Furthermore, the monoamine systems, which are an important component of the substrate for associative learning, are directly affected by antidepressant medication. When this is considered in the context of the clinical features of depressive illness, it suggests that depressive illness may comprise a disorder of associative learning (chapter 2).

Whilst there is extensive animal and human lesion work which suggests that the ventromedial prefrontal cortex is associated with emotional experience, and some imaging studies appear consistent, many others are not. However, a meta-analysis of imaging studies of healthy subjects (chapter 3) was clearly consistent with these non-imaging studies. Notably though, the subgenual anterior cingulate and medial orbitofrontal cortex was not often reported maximally active. In the case of the former, this may be consistent with the hypothesis that the same regions which represent the rewarding and aversive aspects of sensory stimuli, are active when humans experience emotion. The stereotactic meta-analysis method is useful for testing hypotheses of functional segregation, estimating centers of functional activity and boundaries between such centers, the former being used in an fMRI study of depressive illness (chapter 5). As mentioned earlier, a further meta-analyses is currently underway with Chris Frith and Paul Burgess (London University) investigating hypothesised segregation of cognitive functions in the anterior cingulate. Additionally, a separate study investigating hypothesised change in the

pattern of cognitive and emotion related brain activity in depressive illness is being undertaken by psychiatry trainees in Aberdeen.

There is extensive experimental evidence of neural predictive error signals in animals: such signals appear to conform to various formal learning theories which include the Kalman filter (chapter 4). Many brain structures exhibit such signals, including the monoamine systems and ventromedial prefrontal cortex. Recently, there have been many reports of the same signals being detected in healthy humans. The hypothesis, that depressive illness consists of abnormal associative learning, was explored with an fMRI study (chapter 5). More specifically it was predicted that, during a game involving unexpected winning and losing, the brain would automatically create an optimal prediction of events, approximated by a Kalman filter, and compare this with the actual cognitive-perceptual signal to create an error signal. It was further hypothesised that patients with a depressive illness would exhibit a pattern of error signals which differed from controls, reflecting an abnormality of associative learning.

The fMRI study found evidence of a Kalman filter derived error signal in control subjects. In the case of patients, the error signal was significantly increased within prespecified brain regions. Within the patient group alone, for the same brain regions, the error signal correlated with depressive illness severity. Additionally, structural equation modelling was used to investigate hypothesised change in effective connectivity of the error signal, between the prespecified regions. A significant difference was found, which in some cases also correlated with illness severity. The findings of this study therefore support the initial hypothesis. This is the first study to investigate predictive error signals in a psychiatric disorder.

Interpretation of the systematically increased error signal in patients is made difficult by the ambiguity of the relationship between the BOLD fMRI signal and underlying neuronal events. Whilst many parts of the neural processing could give rise to an increased error signal, it seems parsimonious to suggest that an increased error signal reflects an increased mismatch between (the model) predictions and the cognitive-perceptual signals corresponding to the actual events. Such a mismatch could be the result of an abnormality of the cognitive-perceptual signal, or more likely, model generation (presumably involving neural plasticity mechanisms

triggered via error signals). Of course though, the comparator operation, which is assumed to generate the error signal, could itself be abnormal.

It has been suggested that the brain learns and maintains an internal model of the external world, and that theories of associative learning may help to understand such models. More speculatively, it has been suggested that activation and modification of such models by incoming information comprises subjective experience or consciousness (chapter 4). If depressive illness does involve abnormal associative learning, manifested by abnormal predictive error signals, this hypothesis might provide a link between the subjective aspects of the illness, cognitive features such as negative automatic thoughts, and the mode of action of antidepressants.

It has been suggested that in unipolar depressive illness, there is a subtle abnormality of the midline of the midbrain which includes the monoamine systems. Since these systems exhibit predictive error signals, which are broadcast to widespread brain regions, including those reported as having abnormal function in studies of depressive illness, such an anatomical abnormality might be a cause of the abnormal error signals in depressed patients. Since no abnormality was found (chapter 6), this can not be the cause of the increased predictive error signals.

In summary, this thesis has described an investigation into the mechanisms of depressive illness. The work has centered on predictive error signals, which have been reported to be present in many animal and human studies. Systematically increased error signals in depressed patients were found. The significance of this finding has been discussed in the context of clinical features of depressive illness, imaging and non-imaging studies, and treatments for depression, including antidepressants, psychotherapy and physical treatments. A number of potential future studies have been suggested. The most important of these are summarised in the last section.

7.2 FUTURE WORK

The most obvious future work is simply to determine whether the fMRI study results can be consistently replicated with other subjects and with different paradigms. In attempting replication, it is important to consider the issue of study power. Whilst random effects analysis allows generalization to the sampled

population, the statistical power is significantly lower than in a fixed effects analysis. Most imaging studies are of low power (small number of subjects) due to the difficulty of doing such studies. Additionally, patients and controls are not randomly sampled but comprise highly selected subjects willing and able to take part in an imaging study. Consequently, attempts at replication with *different* groups of subjects using *fixed effects* analyses may be best, particularly for minimizing type 2 error. Additionally, detection of the error signal, and any abnormalities in depressed patients, will be maximized when the theoretical error signal most closely matches the actual neural error signal. This will require more detailed modeling of the error signal, together with the associated haemodynamic response function. Since signal loss in fMRI tends to be located in or near *a priori* defined regions of interest, methods to quantify and prevent the susceptibility artifact signal loss need to be implemented. A further fMRI study at Aberdeen University is planned which will explore these issues with a different group of patients and a different paradigm.

It is generally accepted that depressed patients characteristically interpret information in a biased manner, which has been described as a tendency to experience negative automatic thoughts. There is considerable evidence that interpretation of all information may require significant "top-down" information processing; this is particularly so when more ambiguous information has to be interpreted. Clearly then, such tendency to biased information processing might be a reflection of an abnormal "model" of self and surroundings: negative automatic thoughts would arise because they are the "most likely" interpretation of the information. Exploring the link between such biased information processing, from a cognitive-behavioral perspective, particularly in the context of different amounts of ambiguity, is of interest. Cognitive-behavioral therapy attempts to treat negative automatic thoughts by first teaching the patient to identify such thoughts, think of alternatives, then challenge the negative automatic thoughts with evidence of contrary views. This might be interpreted as attempting to alter a patient's "model" of themselves, by reducing ambiguity in particular situations, through focused contrary evidence. Tests of this hypothesis are possible; e.g., in unmedicated depressed patients scanned before and after (or during a course of) cognitive

behavioral therapy, is there initially an increased error signal which is subsequently reduced, and to what extent is the effect of such treatment generalised ?

There is robust evidence of predictive error signals in both animals and humans and the monoamine systems exhibit such signals. The common mode of action of antidepressant medication is to increase monoamine levels; however, the effects of antidepressants on the error signals are unknown. Investigating this issue is of particular interest. In such studies it will be important not to confuse short term administration with long term administration, since *opposite* effects may occur (Daw, Kakade & Dayan, 2002) related to the characteristic delayed response to antidepressant treatment (chapter 5). Parallel animal and human studies might be done. A proposal for an investigation into the effects of antidepressant medication on predictive error signals in healthy volunteers has been awarded funding for a PhD student at Aberdeen University. If it is also possible to replicate the fMRI findings in patients, exploring the link between the error signals and neural plasticity, in relation to stressors and antidepressant medication, is likely to be important.

Finally, abnormal error signals may be reported in the future for other psychiatric disorders. For example, abnormalities of associative learning (disrupted latent inhibition and blocking) have been reported in acutely ill unmedicated patients with schizophrenia experiencing positive symptoms (Escobar, Oberling & Miller, 2002), suggesting that this illness might also be associated with abnormal predictive error signals. A common action of antipsychotics is on the mesolimbic dopamine system, and this region represents predictive error signals. Furthermore, Friston has discussed a theory of schizophrenia in which functional disintegration is caused by abnormal plasticity in aminergic systems responsible for associative learning (Friston, 1998). Consequently, investigations into the effects of antipsychotics on dopaminergic predictive error signals, and imaging studies investigating a potential abnormality of such signals in patients with schizophrenia, are also of considerable interest.

REFERENCES

Aggleton, J. P. (2000) *The Amygdala: A Functional Analysis*. New York: Oxford University Press.

Akitsuki, Y., Sugiura, M., Watanabe, J., *et al* (2003) Context-dependent cortical activation in response to financial reward and penalty: an event-related fMRI study. *Neuroimage*, **19**, 1674-1685.

Alexander, G. E. & Crutcher, M. D. (1990) Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *TINS*, **13**, 266-271.

Alexander, G. E., Crutcher, M. D. & DeLong, M. R. (1990) Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. *Prog Brain Res*, **85**, 119-146.

Alexander, G. E., DeLong, M. R. & Strick, P. L. (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, **9**.

Altman, D. G. (1991) *Practical Statistics for Medical Research*. London: Chapman and Hall.

Arbuthnott, G. (1998) Neuropharmacology. In *Companion to Psychiatric Studies* (eds E. C. Johnstone, C. P. L. Freeman & A. K. Zealley). Edinburgh: Churchill Livingstone.

Armony, J. L. & LeDoux, J. E. (2000) How Danger is Encoded: Toward a Systems, Cellular, and Computational Understanding of Cognitive-Emotional Interactions in Fear. In *The New Cognitive Neurosciences* (ed M. S. Gazzaniga). London: MIT Press.

Baddeley, R. J., Ingram, H. A. & Miall, R. C. (2003) System identification applied to a visuomotor task: near-optimal human performance in a noisy changing task. *J Neurosci*, **23**, 3066-3075.

Baev, K. V. (1998) *Biological Neural Networks: The Hierarchial Concept of Brain Function*. Boston: Birkhauser.

Baev, K. V., Greene, K. A., Marciano, F. F., *et al* (2002) Physiology and pathophysiology of cortico-basal ganglia-thalamocortical loops: theoretical and practical aspects. *Prog Neuropsychopharmacol Biol Psychiatry*, **26**, 771-804.

---- (1995) Novel heuristics of functional neural networks: Implications for future strategies in functional neurosurgery. *Stereotact Funct Neurosurg*, **65**, 26-36.

Ballantyne, H. T., Bouckoms, A. J., Thomas, E. K., *et al* (1987) Treatment of psychiatric illness by stereotactic cingulotomy. *Biological Psychiatry*, **22**, 807-819.

Bandler, R. & Keay, K. A. (1996) Columnar organization in the midbrain periaqueductal gray and the integration of emotional expression. *Prog Brain Res*, **107**, 285-300.

Bardo, M. T. (1998) Neuropharmacological mechanisms of drug reward: beyond dopamine in the nucleus accumbens. *Clinical Reviews in Neurobiology*, **12**, 37-67.

Barlow, H. B. (1985) Cerebral cortex as model builder. In *Models of the visual cortex*. New York: Wiley.

Bartels, A. & Zeki, S. (1998) The theory of multistage integration in the visual brain. *Proc R Soc Lond B Biol Sci*, **265**, 2327-2332.

Basser, P. J. (1995) Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR Biomed*, **8**, 333-344.

Basser, P. J., Mattiello, J. & LeBihan, D. (1994) Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B*, **103**, 247-254.

Basser, P. J. & Pierpaoli, C. (1998) A simplified method to measure the diffusion tensor from seven MR images. *Magn Reson Med*, **39**, 928-934.

Bastin, M. E. & Armitage, P. A. (2000) On the use of water phantom images to calibrate and correct eddy current induced artefacts in MR diffusion tensor imaging. *Magn Reson Imaging*, **18**, 681-687.

Beale, R. (1990) *Neural Computing: An Introduction*. Bristol: Adam Hilger.

Beck, A. T., Ward, H. C. & Mendelson, M. (1961) An inventory for measuring depression. *Arch Gen Psychiatry*, **4**, 561-571.

Becker, G., Becker, T., Struck, M., *et al* (1995) Reduced echogenicity of brainstem raphe specific to unipolar depression: A transcranial color-coded real-time sonography study. *Biological Psychiatry*, **38**, 180-184.

Becker, G., Berg, D., Lesch, K. P., *et al* (2001) Basal limbic system alteration in major depression: a hypothesis supported by transcranial sonography and MRI findings. *International Journal of Neuropsychopharmacology*, **4**, 21-31.

Becker, G. & Griewing, B. (1998) Examination techniques. In *Echoenhancers and Transcranial Color Duplex Sonography* (eds U. Bogdahn, G. Becker & F. Schlachetzki). Berlin, Vienna: Blackwell.

Becker, G., Struck, M., Bogdahn, U., *et al* (1994a) Echogenicity of the Brainstem in Patients with Major Depression. *Psychiatry Research: Neuroimaging*, **55**, 75-84.

---- (1994b) Echogenicity of the brainstem raphe in patients with major depression. *Psychiatry Res*, **55**, 75-84.

Becker, J. T., Mintun, M. A., Diehl, D. J., *et al* (1994) Functional neuroanatomy of verbal free recall: a replication study. *Human Brain Mapping*, **1**, 284-292.

Becker, T., Becker, D., Berg, D., *et al* (1998) Pathological findings in neuropsychiatric diseases. In *Echoenhancers and Transcranial Color Duplex Sonography* (eds U. Bogdahn, G. Becker & F. Schlachetzki). London: Blackwell Science.

Becker, T., Becker, G., Seufert, J., *et al* (1997) Parkinson's disease and depression: evidence for an alteration in the basal limbic system detected by transcranial sonography. *Journal of Neurology, Neurosurgery, and Psychiatry*, **63**, 590-596.

Berg, D., Supprian, T., Hoffmann, E., *et al* (1999) Depression in Parkinson's disease: brainstem midline alteration on transcranial sonography and magnetic resonance imaging. *Journal of Neurology*, **246**, 1186-1193.

Berg, D., Supprian, T., Thomae, J., *et al* (2000) Lesion pattern in patients with multiple sclerosis and depression. *Mult Scler*, **6**, 156-162.

Berns, G. S., McClure, S. M., Pagnoni, G., *et al* (2001) Predictability modulates human brain response to reward. *J Neurosci*, **21**, 2793-2798.

Bhagwagar, Z., Cowen, P. J., Goodwin, G. M., *et al* (2004) Normalization of enhanced fear recognition by acute SSRI treatment in subjects with a previous history of depression. *Am J Psychiatry*, **161**, 166-168.

Bloom, F. E. & Kupfer, D. J. (eds) (1995) *Psychopharmacology: the fourth generation of progress*. New York: Raven Press.

BNF (2004) *British National Formulary*. London: Pharmaceutical Press.

Bonhomme, N. & Esposito, E. (1998) Involvement of serotonin and dopamine in the mechanism of action of novel antidepressant drugs: A review. *Journal of Clinical Psychopharmacology*, **18**, 447-454.

Borah, J., Young, L. R. & Curry, R. E. (1988) Optimal estimator model for human spatial orientation. *Ann N Y Acad Sci*, **545**, 51-73.

Bosquet, O., Balakrishnon, K. & Honavar, V. (1999) Is the hippocampus a Kalman filter. *Proceedings of the Pacific Symposium on Biocomputing*, **3**, 619-630.

Bush, G., Luu, P. & Posner, M. I. (2000) Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn Sci*, **4**, 215-222.

Bush, G., Vogt, B. A., Holmes, J., *et al* (2002) Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proc Natl Acad Sci U S A*, **99**, 523-528.

Bush, G., Whalen, P. J., Rosen, B. R., *et al* (1998) The counting Stroop: An interference task specialised for functional neuroimaging- Validation study with functional MRI. *Human Bran Mapping*, **6**, 270-282.

Cabeza, R. & Nyberg, L. (2000) Imaging cognition II: An empirical review of 275 PET and fMRI studies. *J Cogn Neurosci*, **12**, 1-47.

Cardinal, R. N., Parkinson, J. A., Hall, J., *et al* (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci Biobehav Rev*, **26**, 321-352.

Carroll, B. J. (1994) Brain mechanisms in manic depression. *Clinical Chemistry*, **40**, 303-308.

Churchland, P. S. & Sejnowski, T. J. (1992) *The Computational Brain*. Cambridge, MA: MIT Press.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences* (2nd edn). Hillsdale, NJ: Erlbaum.

CRAG Working Group (1996) *Neurosurgery for Mental Disorder*: HMSO Scotland (J2318 7/96).

Critchley, H. D., Elliott, R., Mathias, C. J., *et al* (2000) Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *The Journal of Neuroscience*, **20**, 3033-3040.

Davis, M., Hitchcock, J. M. & Rosen, J. B. (1987) Anxiety and the amygdala: pharmacological and anatomical analysis of fear potentiated startle. In *The psychology of learning and motivation* (ed G. H. Bower). San Diego: Academic Press.

Davis, M., Rainnie, D. & Cassell, M. (1994) Neurotransmission in the rat amygdala related to fear and anxiety. *Trends Neurosci*, **17**, 208-214.

Daw, N. D., Kakade, S. & Dayan, P. (2002) Opponent interactions between serotonin and dopamine. *Neural Netw*, **15**, 603-616.

Day, B. L., Dressler, D., Maertens de Noordhout, A., *et al* (1989) Electric and magnetic stimulation of human motor cortex: surface EMG and single motor unit responses. *J Physiol*, **412**, 449-473.

Dayan, P., Kakade, S. & Montague, P. R. (2000) Learning and selective attention. *Nat Neurosci*, **3 Suppl**, 1218-1223.

Deakin, J. F. W. (1996) 5HT, antidepressant drugs and the psychosocial origins of depression. *Journal of Psychopharmacology*, **10**, 31-38.

Deakin, J. F. W. & Graeff, F. G. (1991) 5-HT and the mechanisms of defence. *Journal of Psychopharmacology*, **5**, 305-315.

DeLong, M. R. (1990) Primate models of movement disorders of basal ganglia origin. *TINS*, **13**, 281-285.

Devinski, O., Morrell, M. & Vogt, B. A. (1995) Contributions of anterior cingulate cortex to behaviour. *Brain*, **118**, 279-306.

Dickinson, A. & Balleine, B. (2002) The role of learning in the operation of motivational systems. In *Stevens' Handbook of Experimental Psychology* (eds H. Pashler & R. Gallistel). New York: Wiley.

Dickinson, A. & Dearing, M. F. (1979) Appetitive-aversive interactions and inhibitory processes. In *Mechanisms of Learning and Motivation* (eds A. Dickinson & R. A. Boakes). Hillsdale, NJ: Erlbaum.

Drevets, W. C. (1999) Prefrontal Cortico-Amygdalar Metabolism in Major Depression. *Annals of the New York Academy of Sciences*, **877**, 614-637.

---- (ed) (2000a) *Functional anatomical abnormalities in limbic and prefrontal cortical structures in major depression.* London: Elsevier Science.

---- (2000b) Neuroimaging studies of mood disorders. *Biol Psychiatry*, **48**, 813-829.

Drevets, W. C., Gautier, C., Price, J. C., *et al* (2001) Amphetamine-induced dopamine release in human ventral striatum correlates with euphoria. *Biological Psychiatry*, **49**, 81-96.

Drevets, W. C., Price, J. L., Bardgett, M. E., *et al* (2002) Glucose metabolism in the amygdala in depression: relationship to diagnostic subtype and plasma cortisol levels. *Pharmacol Biochem Behav*, **71**, 431-447.

Drevets, W. C., Price, J. L., Simpson, J. R., Jr., *et al* (1997) Subgenual prefrontal cortex abnormalities in mood disorders. *Nature*, **386**, 824-827.

Drevets, W. C. & Raichle, M. E. (1998) Reciprocal suppression of regional cerebral blood flow during emotional versus higher cognitive processes: implications for interactions between emotion and cognition. *Cognition and Emotion*, **12**, 353-385.

Drevets, W. C. & Todd, R. D. (1997) Depression, Mania and Related Disorders. In *Adult Psychiatry* (ed S. B. Guze). St Louis, MO: Mosby Press.

Dum, R. P. & Strick, P. L. (1993) Cingulate motor areas. In *Neurobiology of cingulate Cortex and Limbic Thalamus* (eds B. A. Vogt & M. Gabriel), pp. 415-441. Boston: Birkhauser.

Duncan, J. & Owen, A. M. (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci*, **23**, 475-483.

Duvernoy, H. M. (1991) *The human brain surface. Three dimensional sectional anatomy and MRI.* Austria: Springer-Verlag.

Ebmeier, K. P., Berge, A., Semple, D., *et al* (2004) Biological Treatments of Mood Disorders. In *Mood Disorders: A Handbook of Science and Practice* (ed M. Power), pp. 158-160. Chichester, UK: Wiley.

Ebmeier, K. P. & Ebert, D. (1996) Imaging functional change and dopaminergic activity in depression. In *Dopamine Disease States* (eds R. J. Beninger & T. Palomo), pp. 511-522. Madrid: CYM Press.

Ebmeier, K. P. & Kronhaus, D. (2002) Brain imaging and mood disorders. In *Biological Psychiatry* (eds H. D'haenen, J. A. den Boer & P. Willner). London: John Wiley and Sons.

Eliasmith, C. & Anderson, C. H. (2002) *Neural Engineering: Computation, Representation and Dynamics in Neurobiological Systems*: MIT Press.

Elkis, H., Friedman, L., Wise, A., *et al* (1995) Meta-analyses of studies of ventricular enlargement and cortical sulcal prominence in mood disorders: comparisons with controls or patients with schizophrenia. *Archives of General Psychiatry*, **52**, 735-746.

Elliott, R., Dolan, R. & Frith, C. D. (2000) Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex*, **10**, 308-317.

Elliott, R., Friston, K. J. & Dolan, R. J. (2000) Dissociable neural responses in human reward systems. *The Journal of Neuroscience*, **20**, 6159-6165.

Elliott, R., Sahakian, B. J., McKay, A. P., *et al* (1996) Neuropsychological impairments in unipolar depression: the influence of perceived failure on subsequent performance. *Psychol Med*, **26**, 975-989.

Elliott, R., Sahakian, B. J., Michael, A., *et al* (1998) Abnormal neural response to feedback on planning and guessing tasks in patients with unipolar depression. *Psychol Med*, **28**, 559-571.

Ellison, G. (2002) Neural degeneration following chronic stimulant abuse reveals a weak link in brain, fasciculus retroflexus, implying the loss of forebrain control circuitry. *Eur Neuropsychopharmacol*, **12**, 287-297.

Endicott, J. & Spitzer, R. L. (1978) A diagnostic interview schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry*, **35**, 837-844.

Erdfelder, E., Faul, F. & Buchner, A. (1996) GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, and Computers*, **28**, 1-11.

Erickson, K., Drevets, W. & Schulkin, J. (2003) Glucocorticoid regulation of diverse cognitive functions in normal and pathological emotional states. *Neurosci Biobehav Rev*, **27**, 233-246.

Escobar, M., Oberling, P. & Miller, R. R. (2002) Associative deficit accounts of disrupted latent inhibition and blocking in schizophrenia. *Neurosci Biobehav Rev*, **26**, 203-216.

Freedman, L. J., Insel, T. R. & Smith, Y. (2000) Subcortical projections of area 25 (subgenual cortex) of the macaque monkey. *J Comp Neurol*, **421**, 172-188.

Freeman, C. P. L., Crossley, D. & Eccelston, D. (eds) (2000) *Neurosurgery for Mental Disorder*. London: Royal College of Psychiatrists.

Friston, K. (2002) Beyond Phrenology: What Can Neuroimaging Tell Us About Distributed Circuitry? *Annu Rev Neurosci*, **25**, 221-250.

Friston, K. J. (1998) The disconnection hypothesis. *Schizophr Res*, **30**, 115-125.

Frith, C. (2003) What do imaging studies tell us about the neural basis of autism? *Novartis Found Symp*, **251**, 149-166; discussion 166-176, 281-197.

Fuster, J. M. (1997) *The Prefrontal Cortex*. New York: Lippincott-Raven.

Garcia, C. E., Prett, D. M. & Morari, M. (1989) Model predictive control: theory and practice - a survey. *Automatica*, **25**, 335-348.

Genovese, C. R., Lazar, N. A. & Nichols, T. (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, **15**, 870-878.

Glasauer, S. & Merfield, D. M. (1997) Modelling three dimensional vestibular responses during complex motion stimulation. In *Three-Dimensional Kinematics of Eye, Head and Limb Movements* (eds M. Fetter, T. Haslwanter & H. Misslisch). Amsterdam: Harwood academic publishers.

Glascher, J. P. & Buchel, C. (2004) Formal learning theory predicts activation in the human amygdala. In *Human Brain Mapping* (eds S. Y. Bookheimer, J.-B. Poline & B. Gulyas). Budapest: Elsevier.

Goldberger, A. L., Amaral, L. A. N., Hausdorff, J. M., *et al* (2002) Fractal dynamics in physiology: Alterations in disease and aging. *Proc Natl Acad Sci U S A*, **99**, 2466-2472.

Good, P. (1994) *Permutation Tests*. London: Springer-Verlag.

Goodwin, G. (1998) Mood Disorder. In *Companion to Psychiatric Studies* (eds E. C. Johnstone, C. P. L. Freeman & A. K. Zealley). Edinburgh: Churchill Livingstone.

Graeff, F. G. (1993) Role of 5-HT in defensive behavior and anxiety. *Rev Neurosci*, **4**, 181-211.

---- (2004) Serotonin, the periaqueductal gray and panic. *Neurosci Biobehav Rev*, **28**, 239-259.

Grafton, S. T., Sutton, J., Couldwell, W., *et al* (1994) Network analysis of motor system connectivity in Parkinson's disease: modulation of thalamocortical interactions after pallidotomy. *Human Brain Mapping*, **2**, 45-55.

Gray, J. A. (1981) Anxiety as a paradigm case of emotion. *British Medical Bulletin*, **37**, 193-197.

---- (1995) The contents of consciousness: a neuropsychological conjecture. *Behav Brain Sci*, **18**, 659-722.

---- (2004) *Consciousness: creeping up on the hard problem*. Oxford: Oxford University Press.

Gray, J. A. & McNaughton, N. (2000) *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System* (2nd edn). Oxford: Oxford University Press.

Gross, J. J. (1998) The emerging field of emotion regulation: an integrative review. *Review of General Psychology*, **2**, 271-299.

Gross, R. (1996) *Psychology: The Science of Mind and Behaviour*. London: Hodder and Stoughton.

Grush, R. (2004) The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*.

Grzywacz, N. M. & de Juan, J. (2003) Sensory adaptation as Kalman filtering: theory and illustration with contrast adaptation. *Network*, **14**, 465-482.

Guyton, A. (1991) *Textbook of Medical Physiology*. London: W.B. Saunders.

Hamilton, M. (1960) Rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry*, **23**, 56-62.

Hardy, R. N. (1981) *Endocrine Physiology*. London: Arnold.

Harmer, C. J., Hill, S. A., Taylor, M. J., *et al* (2003) Toward a neuropsychological theory of antidepressant drug action: increase in positive emotional bias after potentiation of norepinephrine activity. *Am J Psychiatry*, **160**, 990-992.

Harmer, C. J., Shelley, N. C., Cowen, P. J., *et al* (2004) Increased positive versus negative affective perception and memory in healthy volunteers following

selective serotonin and norepinephrine reuptake inhibition. *Am J Psychiatry*, **161**, 1256-1263.

Harrison-Read, P., Tyrer, P. & Sharpe, M. (2004) Neurotic, stress-related and somatoform disorders. In *Companion to Psychiatric Studies* (eds E. C. Johnstone, D. G. Owens, S. M. Lawrie, *et al*). Edinburgh: Churchill Livingstone.

Hawkins, R. D. & Kandel, E. R. (1984) Is there a cell based biological alphabet for simple forms of learning ? *Psychological Review*, **91**, 375-391.

Hawton, K., Salkoviskis, P. M., Kirk, J., *et al* (1994) *Cognitive Behaviour Therapy for Psychiatric Problems: A Practical Guide*. Oxford: Oxford University Press.

Haxby, J. V., Gobbini, M. I., Furey, M. L., *et al* (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, **293**, 2425-2430.

Heath, R. G. (1954) Definition of the Septal Region. In *Studies in Schizophrenia*. London: Oxford University Press.

---- (1964) Pleasure response of human subjects to direct stimulation of the brain: physiologic and psychodynamic considerations. In *The Role of Pleasure in Behaviour* (ed R. G. Heath), pp. 219-243. New York: Hober.

Hebb, D. O. (1949) *The Organisation of Behaviour*. New York: Wiley.

Higgins, G. A. & Fletcher, P. J. (2003) Serotonin and drug reward: focus on 5-HT2C receptors. *Eur J Pharmacol*, **480**, 151-162.

Hobson, J. A., Pace-Schott, E. F. & Stickgold, R. (2000) Dreaming and the brain: toward a cognitive neuroscience of conscious states. *Behav Brain Sci*, **23**, 793-1121.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., *et al* (2003) Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport*, **14**, 2481-2484.

Hung, G. K. (2001) *Models of Oculomotor Control*: World Scientific Pub Co.

Ingber, A. L. (1989) Very fast simulated re-annealing. *Mathematical computer modelling*, **12**, 967-973.

Inghilleri, M., Berardelli, A., Cruccu, G., *et al* (1993) Silent period evoked by transcranial stimulation of the human cortex and cervicomedullary junction. *J Physiol*, **466**, 521-534.

Ishai, A., Ungerleider, L. G., Martin, A., *et al* (1999) Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A*, **96**, 9379-9384.

Jeffery, K. J. & Reid, I. C. (1997) Modifiable neuronal connections: an overview for psychiatrists. *Am J Psychiatry*, **154**, 156-164.

Jeste, D. V., Lohr, J. B. & Goodwin, F. K. (1988) Neuroanatomical studies of major affective disorders: a review and suggestions for further research. *British Journal of Psychiatry*, **153**, 444-459.

Job, D. E., Whalley, H. C., McConnell, S., *et al* (2002) Structural gray matter differences between first-episode schizophrenics and normal controls using voxel-based morphometry. *Neuroimage*, **17**, 880-889.

Johnstone, E. C. (1998) Psychiatry - its history and boundaries. In *Companion to Psychiatric Studies* (eds E. C. Johnstone, C. P. L. Freeman & A. K. Zealley). Edinburgh: Churchill Livingstone.

Kalman, R. E. (1960) A new approach to linear prediction and control problems. *Transactions of the ASME: Journal of Basic Engineering*, **82**, 35-45.

Kamin, L. J. (1969) Selective associations and conditioning. In *Fundamental Issues in Instrumental Learning* (eds N. J. Mackintosh & W. K. Honig). Halifax, Cannada: Dalhousie University Press.

Kandel, E. R., Schwartz, J. H. & Jessell, T. M. (eds) (2000) *Principles of Neural Science*. New York: McGraw-Hill.

Kendell, R. E., Lawrie, S. M. & Johnstone, E. C. (2004) Diagnosis and classification. In *Companion to Psychiatric Studies* (eds E. C. Johnstone, D. G. Cunningham Owens, S. M. Lawrie, *et al*), pp. 243-255. Edinburgh: Churchill Livingstone.

Khoo, M. C. K. (2000) *Physiological Control Systems: Analysis, Simulation and Estimation*. Piscataway, NJ: IEEE Press.

Klerman, G. L. & Weissman, M. M. (1984) *Interpersonal Psychotherapy of Depression*. New York: Basic Books.

Knapp, T. R. (1978) Canonical correlation analysis: a general parametric significance testing system. *Psychological Bulletin*, **85**, 410-416.

Knight, G. (1964) The orbital cortex as an objective in the surgical treatment of mental illness. *British Journal of Surgery*, **51**, 114-124.

Konorski, J. (1967) *Integrative activity of the brain: an interdisciplinary approach.* Chicago: University of Chicago Press.

Koob, G. F. & Bloom, F. E. (1988) Cellular and molecular mechanisms of drug dependence. *Science*, **242**, 715-723.

Kril, J. J. & Halliday, G. M. (1999) Brain shrinkage in alcoholics: a decade on and what have we learned? *Progress in Neurobiology*, **58**, 381-387.

Kringelbach, M. L. & Rolls, E. T. (2004) The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Prog Neurobiol*, **72**, 341-372.

Kropotov, J. D. & Etlinger, S. C. (1999) Selection of actions in the basal ganglia-thalamocortical circuits: review and model. *International Journal of Psychophysiology*, **31**.

Lang, J. (1993) Surgical anatomy of the brain stem. *Neurosurg Clin N Am*, **4**, 367-403.

Lang, P. J., Bradley, M. M. & Cuthbert, B. N. (1998) Emotion, Motivation and Anxiety: Brain Mechanisms and Psychophysiology. *Biological Psychiatry*, **44**, 1248-1263.

Lavric, A. & Wills, A. J. (2004) Distinguishing between formal theories of associative learning. In *Human Brain Mapping* (eds S. Y. Bookheimer, J.-B. Poline & B. Gulyas). Budapest: Elsevier.

Lawley, D. N. & Maxwell, A. E. (1971) *Factor analysis as a statistical method.* London: Butterworths.

Lawrie, S. M. & Abukmeil, S. S. (1998) Brain abnormality in schizophrenia. A systematic and quantitative review of volumetric magnetic resonance imaging studies. *Br J Psychiatry*, **172**, 110-120.

Lawrie, S. M., McIntosh, A. M. & Rao, S. (2000) *Critical Appraisal for Psychiatry*. Edinburgh: Churchill Livingstone.

LeDoux, J. (1998a) *The Emotional Brain*. London: Phoenix.

---- **(1998b)** Fear and the brain: where have we been, and where are we going? *Biological Psychiatry*, **44**, 1229-1238.

Liotti, M., Mayberg, H. S., Brannan, S. K., *et al* (2000) Differential limbic-cortical correlates of sadness and anxiety in healthy subjects: Implications for affective disorders. *Biological Psychiatry*, **48**, 30-42.

Lipschutz, B., Friston, K. J., Ashburner, J., *et al* (2001) Technical Note. Assessing study-specific regional variations in fMRI signal. *NeuroImage*, **13**, 392-398.

Mai, J. K., Assheuer, J. & Paxinos, G. (1998) *Atlas of the Human Brain*. San Diego: Academic Press.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., *et al* (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*, **19**, 1233-1239.

Marr, D. (1969) A theory of cerebellar cortex. *Journal of Physiology*, **202**, 437-470.

Mayberg, H. S. (2001) Frontal Lobe Dysfunction in Secondary Depression. In *The Frontal Lobes and Neuropsychiatric Illness* (eds S. P. Salloway, P. F. Malloy & J. D. Duffy), pp. 167-186. Washington: American Psychiatric Publishing Inc.

---- **(2003)** Modulating dysfunctional limbic-cortical circuits in depression: towards development of brain-based algorithms for diagnosis and optimised treatment. In *Imaging Neuroscience: Clinical Frontiers for Diagnosis and Management* (eds R. S. Frackowiak & T. Jones), pp. 193-207. Oxford: Oxford University Press.

McArdle, J. J. & McDonald, R. P. (1984) Some algebraic properties of the Reticular Action Model for moment structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 234-251.

McBride, W. J., Murphy, J. M. & Ikemoto, S. (1999) Localization of brain reinforcement mechanisms: intracranial self-administration and intracranial place-conditioning studies. *Behavioural Brain Research*, **101**, 129-152.

McClure, S. M., Berns, G. S. & Montague, P. R. (2003) Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, **38**, 339-346.

McClure, S. M., Li, J., Cohen, J., *et al* (2004) Neural correlates of reinforcement learning in a dynamic economic game. In *Human Brain Mapping* (eds S. Y. Bookheimer, J.-B. Poline & B. Gulyas). Budapest: Elsevier.

McEwan, B. S. (2003) Mood disorders and allostatic load. *Biological Psychiatry*, **54**, 200-207.

McIntosh, A. R. (2000) Towards a network theory of cognition. *Neural Netw*, **13**, 861-870.

McIntosh, C. & Ritson, B. (2001) Treating depression complicated by substance misuse. *Advances in Psychiatric Treatment*, **7**, 357-364.

McLaren, I. (1989) The computational unit as an assembly of neurones: an implementation of an error correcting learning algorithm. In *The Computing Neuron* (eds R. Durbin, C. Miall & G. Mitchison). Amsterdam: Addison-Wesley.

McLean, P. D. (1949) Psychosomatic disease and the 'visceral brain'. Recent developments bearing on the Papez theory of emotion. *Psychosomatic Medicine*, **11**, 338-353.

Mega, M. S. & Cummings, J. L. (1997) The Cingulate and Cingulate Syndromes. In *Contempory Behavioural Neurology* (eds M. R. Trimble & J. L. Cummings), pp. 189-214. Oxford: Butterworth-Heinemann.

Mercado, E., 3rd, Myers, C. E. & Gluck, M. A. (2001) A computational model of mechanisms controlling experience-dependent reorganization of representational maps in auditory cortex. *Cogn Affect Behav Neurosci*, **1**, 37-55.

Meyerson, B. A. & Mindus, P. (1988) The role of the anterior internal capsulotomy in psychiatric surgery. In *Modern Stereotactic Neurosurgery* (ed L. D. Lunsford), pp. 353-363. Lancaster: Martinus Nijhoff Publishing.

Millenson, J. R. (1967) *Principles of behavioural analysis*. New York: MacMillan.

Montgomery, S. A. & Asberg, M. (1979) A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, **134**, 382-389.

Morris, J. S., Smith, K. A., Cowen, P. J., *et al* (1999) Covariation of Activity in Habenula and Dorsal Raphae Nuclei Following Tryptophan Depletion. *NeuroImage*, **10**, 163-172.

Naumann, M., Becker, G., Toyka, K. V., *et al* (1996) Lenticular nucleus lesion in idiopathic dystonia detected by transcranial sonography. *Neurology*, **47**, 1284-1290.

Neale, M. C. (1997) Mx: statistical modeling. Box 126 MCV, Richmond VA 23298: Department of Psychiatry, Virginia Commonwealth University.

Nelson, H. E. & Wilson, J. R. (1991) *The revised national adult reading test - test manual*. Winsor: NFER-Wilson.

Northrop, R. B. (2000) *Endogenous and Exogenous Regulation and Control of Physiological Systems*. Florida: Chapman and Hall.

Nuttin, B., Gybels, J. & Meyerson, B. (1999) Electrical stimulation in anterior limbs of internal capsules of patients with obsessive compulsive disorder. *The Lancet*, **354**, 1526.

O'Carroll, R. (2004) Neuropsychology. In *Companion to Psychiatric Studies* (eds E. C. Johnstone, D. G. Owens, S. M. Lawrie, *et al*). Edinburgh: Churchill Livingstone.

O'Doherty, J. P., Dayan, P., Friston, K., *et al* (2003) Temporal difference models and reward-related learning in the human brain. *Neuron*, **38**, 329-337.

O'Keefe, J. & Nadel, L. (1978) *The hippocampus as a cognitive map*. Oxford: Clarendon Press.

Olds, J. & Olds, M. E. (1964) The mechanisms of voluntary behaviour. In *Role of Pleasure in Behaviour* (ed R. G. Heath). New York: Harper and Row.

Ongur, D. & Price, J. L. (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex*, **10**, 206-219.

Overall, J. E. & Gorham, D. R. (1962) The brief psychiatric rating scale. *Psychological Reports*, **10**, 799-812.

Owens, D. G., McKenna, P. J. & Davenport, R. (2004) Clinical assessment: interviewing and examination. In *Companion to Psychiatric Studies* (eds E. C. Johnstone, D. G. Owens, S. M. Lawrie, *et al*). Edinburgh: Churchill Livingstone.

Owens, D. G. C. (1998) Clinical Psychopharmacology. In *Companion to Psychiatric Studies* (eds E. C. Johnstone, C. P. L. Freeman & A. K. Zealley). Edinburgh: Churchill Livingstone.

Pagnoni, G., Zink, C. F., Montague, P. R., *et al* (2002) Activity in human ventral striatum locked to errors of reward prediction. *Nat Neurosci*, **5**, 97-98.

Paulin, M. (1988) A Kalman filter theory of the cerebellum. In *Dynamic Interactions in Neural Networks: Models and Data* (eds M. A. Arbib & S. Amari). London: Springer-Verlag.

Pearce, J. M. & Hall, G. (1980) A model for Pavlovian conditioning: variations in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review*, **87**, 532-552.

Peng, C.-K., Hausdorff, J. M. & Goldberger, A. L. (2000) Fractal mechanisms in neural control: Human heartbeat and gait dynamics in health and disease. In *Self-Organised Biological Dynamics and Nonlinear Control* (ed J. Walleczek). Cambridge: Cambridge University Press.

Peyron, R., Laurent, B. & Garcia-Larrea, L. (2000) Functional imaging of brain responses to pain. A review and meta-analysis (2000). *Neurophysiol Clin*, **30**, 263-288.

Phan, K. L., Wagner, T., Taylor, S. F., *et al* (2002) Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, **16**, 331-348.

Picton, T. W. & Stuss, D. T. (1994) Neurobiology of conscious experience. *Curr Opin Neurobiol*, **4**, 256-265.

Ploghaus, A., Tracey, I., Clare, S., *et al* (2000) Learning about pain: the neural substrate of the prediction error for aversive events. *Proc Natl Acad Sci U S A*, **97**, 9281-9286.

Powers, W. T. (1981) *Behavior: The Control of Perception*. Chicago: Aldine Publishing Company.

Press, W. H., Teukoloski, S. A., Vetterling, W. T., *et al* (1994) *Numerical Recipies in C: The Art of Scientific Computing*. Cambridge: Cambridge University Press.

Price, J. L. (1999) Prefrontal cortical networks related to visceral function and mood. *Annals of New York Academy of Sciences*, **877**, 383-396.

Priori, A., Berardelli, A., Mercuri, B., *et al* (1995) The effect of hyperventilation on motor cortical inhibition in humans: a study of the electromyographic silent period evoked by transcranial brain stimulation. *Electroencephalogr Clin Neurophysiol*, **97**, 69-72.

Rao, R. P. (1999) An optimal estimation approach to visual perception and learning. *Vision Res*, **39**, 1963-1989.

Rao, R. P. & Ballard, D. H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*, **2**, 79-87.

Reid, I. C. & Stewart, C. A. (2001) How antidepressants work: new perspectives on the pathophysiology of depressive disorder. *Br J Psychiatry*, **178**, 299-303.

Rescorla, R. A. & Wagner, A. R. (1972) A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory* (eds A. H. Black & W. F. Prokasy). New York: Appleton Century Crofts.

Richardson, A. (1973) Stereotactic limbic leucotomy: surgical technique. *Postgraduate Medical Journal*, **49**, 860-864.

Robbins, T. W. & Everitt, B. J. (1995) Arousal systems and attention. In *The Cognitive Neuroscieces* (ed M. S. Gazzaniga). Cambridge MA: MIT Press.

Rolls, E. T. (1999) *The Brain and Emotion*. Oxford: Oxford University Press.

Rolls, E. T., Inoue, K. & Browning, A. (2003) Activity of primate subgenual cingulate cortex neurons is related to sleep. *J Neurophysiol*, **90**, 134-142.

Rolls, E. T. & Treves, A. (2001) *Neural Networks and Brain Function*. Oxford: Oxford University Press.

Rousseau, P. J. (1987) *Robust regression and Outlier Detection*. New York: Wiley-Interscience.

Samii, A., Wassermann, E. M., Ikoma, K., *et al* (1996) Decreased postexercise facilitation of motor evoked potentials in patients with chronic fatigue syndrome or depression. *Neurology*, **47**, 1410-1414.

Sanfilipo, M., Lafargue, T., Rusinek, H., *et al* (2000) Volumetric measure of the frontal and temporal lobe regions in schizophrenia: relationship to negative symptoms. *Arch Gen Psychiatry*, **57**, 471-480.

Schildkraut, J. J. (1965) The catecholamine hypothesis of mood disorders: a review of supporting evidence. *Am J Psychiatry*, **122**, 509-522.

Schmahmann, J. D. (1998) A new role for the cerebellum: the modulation of cognition and affect. In *Movement Disorders in Neurology and Psychiatry* (eds A. B. Joseph & R. R. Young). Oxford: Blackwell Science Ltd.

Schultz, W., Dayan, P. & Montague, P. R. (1997) A neural substrate of prediction and reward. *Science*, **275**, 1593-1599.

Schultz, W. & Dickinson, A. (2000) Neuronal coding of prediction errors. *Annu Rev Neurosci*, **23**, 473-500.

Schultz, W., Tremblay, L. & Hollerman, J. R. (1998) Reward prediction in primate basal ganglia and frontal cortex. *Neuropharmacology*, **37**, 421-429.

Selye, H. (1950) *The physiology and pathophysiology of exposure to stress*. Montreal: Acta.

Seymour, B., O'Doherty, J. P., Dayan, P., *et al* (2004a) Temporal difference models describe higher-order learning in humans. *Nature*, **429**, 664-667.

Seymour, B. J., O'Doherty, J. P., Dayan, P., *et al* (2004b) The computational basis for higher order pain perception. In *Human Brain Mapping* (eds S. Y. Bookheimer, J.-B. Poline & B. Gulyas). Budapest: Elsevier.

Shah, P. J., Ebmeier, K. P., Glabus, M. F., *et al* (1998) Cortical grey matter reductions associated with treament-resistant unipolar depression. Controlled magnetic resonance imaging study. *British Journal of Psychiatry*, **172**, 527-532.

Shajahan, P. M., Glabus, M. F., Gooding, P. A., *et al* (1999) Reduced cortical excitability in depression. Impaired post-exercise motor facilitation with transcranial magnetic stimulation. *Br J Psychiatry*, **174**, 449-454.

Shajahan, P. M., Glabus, M. F., Steele, J. D., *et al* (2002) Left dorso-lateral repetitive transcranial magnetic stimulation affects cortical excitability and functional connectivity, but does not impair cognition in major depression. *Prog Neuropsychopharmacol Biol Psychiatry*, **26**, 945-954.

Shannon, C. E. (1948) A mathematical theory of communication. *The Bell System Technical Journal*, **27**, 623-656.

Sheline, Y. I., Wang, P. W., Gado, M. H., *et al* (1996) Hippocampal atrophy in recurrent major depression. *Proceedings of the National Academy of Sciences*, **93**, 3908-3913.

Siegle, G. J. (ed) (1999) *A neural network model of attention biases in depression.* New York: Elsevier Science.

Sims, A. (1995) *Symptoms in the Mind.* London: W.B. Saunders Company.

Sirosh, J. & Miikkulainen, R. (1997) Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Comput*, **9**, 577-594.

Solomon, R. L. (1977) An opponent-process theory of acquired motivation: the affective dynamics of addiction. In *Psychopathology: experimental models* (eds J. D. Maser & M. E. P. Seligman). San Francisco: W.H. Feeeman.

---- (1980a) The opponent-process theory of acquired motivation. *American Psychologist*, **35**, 691-712.

---- (1980b) Recent experiments testing an opponent-process theory of acquired motivation. *Acta Neurobiol Exp*, **40**, 271-289.

Solomon, R. L. & Corbit, J. D. (1973) An opponent-process theory of motivation: II Cigarette addiction. *Journal of Abnormal Psychology*, **81**, 158-171.

---- (1974) An opponent-process theory of motivation: 1 Temporal dynamics of affect. *Psychological Review*, **81**, 119-145.

Steele, J. D., Glabus, M. F., Shajahan, P. M., *et al* (2000) Increased cortical inhibition in depression: a prolonged silent period with transcranial magnetic stimulation (TMS). *Psychol Med*, **30**, 565-570.

Steele, J. D. & Lawrie, S. M. (2004a) Neuroimaging. In *Edinburgh Companion to Psychiatric Studies* (eds E. C. Johnstone & S. M. Lawrie). Edinburgh: Churchill Livingstone.

---- (2004b) Segregation of cognitive and emotional function in the prefrontal cortex: a stereotactic meta-analysis. *Neuroimage*, **21**, 868-875.

Steele, J. D., Meyer, M. & Ebmeier, K. P. (2004) Neural Predictive Error Signal Correlates with Depressive Illness Severity in a Game Paradigm. *NeuroImage*, **23**, 269-280.

Strange, B. A., Fletcher, P. C., Henson, R. N., *et al* (1999) Segregating the functions of human hippocampus. *Proc Natl Acad Sci U S A*, **96**, 4034-4039.

Stroup, D. F., Berlin, J. A., Morton, S. C., *et al* (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *Jama*, **283**, 2008-2012.

Stuss, D. T., Picton, T. W. & Alexander, M. P. (2001) Consciousness, self-awareness, and the frontal lobes. In *The Frontal Lobes and Neuropsychiatric Illness* (eds S. P. Salloway, P. F. Malloy & J. D. Duffy). London: American Psychiatric Publishing Inc.

Suri, R. E. (2001) Anticipatory responses of dopamine neurons and cortical neurons reproduced by internal model. *Exp Brain Res*, **140**, 234-240.

---- (2002) TD models of reward predictive responses in dopamine neurons. *Neural Netw*, **15**, 523-533.

Sutton, R. S. & Barto, A. G. (1998) *Reinforcement Learning.* Cambridge, MA: MIT Press.

Tabachnick, B. G. & Fidell, L. S. (1996) *Using multivariate statistics*. New York: HarperCollins Publishers Inc.

Talairach, J., Bancaud, J., Geier, S., *et al* (1973) The cingulate gyrus and human behaviour. *Electroencephalography and Clinical Neurophysiology*, **34**, 45-52.

Talairach, J. & Tournoux, P. (1988) *Co-Planar Stereotaxic Atlas of the Human Brain*. Stuttgart: Thieme Medical Publishers.

Tasker, R. R., Yamashiro, K., Lenz, F., *et al* (1988) Thalmotomy for Parkinsons disease: microelectrode technique. In *Modern Stereotactic Neurosurgery* (ed L. D. Lunsford), pp. 297-314. Lancaster: Martinus Nijhoff Publishing.

Terajima, K. & Nakada, T. (2002) EZ-tracing: a new ready-to-use algorithm for magnetic resonance tractography. *J Neurosci Methods*, **116**, 147-155.

Tremblay, L. K., Naranjo, C. A., Cardenas, L., *et al* (2002) Probing brain reward system function in major depressive disorder: altered response to dextroamphetamine. *Arch Gen Psychiatry*, **59**, 409-416.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., *et al* (2002) Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage*, **16**, 765-780.

Videbech, P. (1997) MRI findings in patients with affective disorder: a meta-analysis. *Acta Psychiatrica Scandinavica*, **96**, 157-168.

Vilensky, J. A. & van Hoesen, G. W. (1981) Corticopontine projections from the cingulate cortex in the rhesus monkey. *Brain Res*, **205**, 391-395.

Vogt, B. A., Finch, D. M. & Olson, C. R. (1992) Functional heterogeneity in cingulate cortex: the anterior executive and posterior evaluative regions. *Cerebral Cortex*, **2**, 435-443.

Vogt, B. A. & Gabriel, M. (1993) *Neurobiology of Cingulate Cortex and Limbic Thalamus*. Boston: Birkhauser.

Vogt, B. A., Nimchinsky, E. A., Vogt, L. J., *et al* (1995) Human cingulate cortex: surface features, flat maps, and cytoarchitecture. *The Journal of Comparative Neurology*, **359**, 490-506.

Wall, P. D. (1993) Pain and the placebo response. In *Experimental and theoretical studies of consciousness*. Chichester, UK: Wiley.

Weizcrantz, L. (1968) Emotion. In *Analysis of Behavioural Change* (ed L. Weiskrantz). New York: Harper and Row.

Wiener, N. (1948) *Cybernetics: or Control and Communication in the Animal and the Machine*. New York: MIT Press.

Williams, P. W. & Warwick, R. (eds) (1980) *Gray's Anatomy*. Edinburgh: Churchill Livingstone.

Wing, J. K., Cooper, J. E. & Sartorius, N. (1974) *The measurement and classification of psychiatric symptoms*. Cambridge: Cambridge University Press.

Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. (1995) An internal model for sensorimotor integration. *Science*, **269**, 1880-1882.

Woyshville, M. J., Lackamp, J. M., Eisengart, J. A., *et al* (1999) On the meaning and measurement of affective instability: clues from chaos theory. *Biol Psychiatry*, **45**, 261-269.

Wright, I. C., Rabe-Hesketh, S., Woodruff, P. W., *et al* (2000) Meta-analysis of regional brain volumes in schizophrenia. *Am J Psychiatry*, **157**, 16-25.

Yates, J. (1985) The content of awareness is a model of the world. *Psychological Review*, **92**, 249-284.

Young, E. A., Haskett, R. F., Murphy-Weinberg, V., *et al* (1991) Loss of glucocorticoid fast feedback in depression. *Arch Gen Psychiatry*, **48**, 693-699.

Zeki, S. (2001) Localization and globalization in conscious vision. *Annu Rev Neurosci*, **24**, 57-86.

Zeman, A. (2002) *Consciousness: A Users Guide*. New Haven and London: Yale University Press.

Zinkgraf, S. (1983) Performing factorial multivariate analysis of variance using canonical correlation analysis. *Educational and Psychological Measurement*, **43**, 63-68.