



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

**HIGH-RESOLUTION IMAGING OF THE HUMAN AMYGDALA:
ROLE OF BASOLATERAL AND CENTROMEDIAL COMPLEXES
IN REWARD PROCESSING AND ASSOCIATIVE LEARNING**

Charlotte Prévost

A thesis submitted for the degree of Doctor of Philosophy
of the University of Dublin, Trinity College Dublin.

2012



Thesis 9793

DECLARATION

I declare that this work has not been submitted previously as an exercise for a degree at this or any other university and that it is entirely my own work. I give my permission for the Trinity College Library to lend or copy this thesis upon request.

Signed



Date

29th of March 2012

ACKNOWLEDGEMENTS

The first half of this PhD was carried out at Trinity College Dublin in Ireland and the second half at the California Institute of Technology in the United States. Living at both places has been a very fulfilling experience, and having the whole lab cross the planet as a big happy family has strengthened the bonds between lab members and has made the past three years very important to me. Moving to Dublin and then to California has been as exciting as it has been challenging, as I was getting further and further away from my loved ones and my beautiful France.

None of this adventure would have been possible without John O'Doherty, my supervisor, and I cannot thank him enough for always having been available, patient, understanding and encouraging. I have been extremely lucky to have worked with such an impressive researcher always full of ideas and never-ending inspiration, and I am sincerely grateful to him for having taught me so much about research and having constantly helped me throughout this PhD.

I also wish to thank Paul Dockree who kindly agreed to be my official administrative supervisor after John moved back to California.

During this PhD, I had the privilege to work with Peter Bossaerts with whom we collaborated on two different projects. For your amazing computational skills and your kindness, merci Peter! Likewise, I would like to thank Mike Tyszka for his precious help working on fMRI sequences at Caltech.

I would also like to thank my other collaborators and friends. I wish to thank Jonathan McCabe who invested a lot of time and effort helping me set up our high-resolution protocol. I am grateful to Ryan Jessup for having helped me on the computational side on two consecutive projects, but most importantly for having been like a big brother to me, always listening and advising. I also would like to thank Daniel McNamee for having helped me and being a friend. Last but not least, I want to thank my dear friend Mimi Liljeholm for her indefatigable listening and her astounding advice, whether it was with regard to research or life. I will surely miss our breaks and girlie conversations!

I am also thankful to the other O'Doherty lab members: Vik, Richard, Jacky, Sang-Wan and more particularly my friends Simon and Tobias. Simon has helped me a lot on most of my projects and has been an amazing RA, as well as friend. I am grateful to the very bright students who have worked with me at Caltech, Bo and Soyoung. Other people have made this time special to me both at the Trinity College Institute of Neuroscience: Sojo, Donal, Teresa, Jeff; and at the California Institute of Technology: Ralph, the Rangel lab, the rotating students and the indescribable Caltech students. I also want to say a big thank you to Paul, Dan, Simon and Tim for proofreading different parts of this thesis.

Very special thanks go to my previous lab in Lyon, France which initiated me to the complex world of research. Thank you Jean-Claude for your support and for having taught me so many things. I also want to thank Guigui, Jéré, Xavier and Elise for their help and friendship.

I would further like to thank my family from the bottom of my heart for letting me go away and trusting in my abilities, but most of all for their endless love. I am truly grateful to my mum for teaching me to never forget what really matters in life and to my dad for teaching me that life is what we make it, hence that everything is possible as long as we work for it. They have given me the will, strength and confidence to follow my dreams. I also want to deeply thank my brothers Matthieu and Vincent for our innumerable fights and reconciliations, and foremost for their support and affection. I also thank Gilou, Laurette, Lise, Marion, Sandra and Valouche for making our family merrier.

Finally, I would like to thank my friends in France for still being there for me despite the years away from them: Roomy, Titou, Marine, Marie, Monette, Zouz, Sabine, Pioup, Baptoo and many more... as well as my friends in California, especially Moss, Kyle and Clément for making my Californian experience special.

Je voudrais conclure en remerciant du fond du coeur mon Bastinou pour son soutien et sa tendresse qui me rappellent chaque jour comme la vie est belle et vaut la peine d'être vécue...

SUMMARY

The amygdala is an inherently heterogeneous aggregation of nuclei. Although the animal literature has revealed some segregation of function within these nuclei, it has not been possible to investigate the role of human amygdalar subregions to date due to limitations in the spatial resolution of fMRI.

To overcome this issue which has largely prevented progress in the understanding of human amygdala function, part of this PhD revolved around developing a high-resolution fMRI (hr-fMRI) protocol, which involved a variety of resolution preserving data analysis procedures specifically optimized for the amygdala (including a decreased voxel size, the unwarping of the functional data, a specific normalization of the amygdala, the forgoing of the application of extrinsic spatial smoothing and the removal of physiological fluctuations). Having considerably increased the spatial resolution of our data (more than 4 times that of standard fMRI), it has become possible to extract reliable signal from the human amygdala's main complexes, namely the basolateral (BLA) and centromedial (CeN) complexes.

In this thesis, we present a body of research illustrating initial hr-fMRI efforts to investigate the functional role of these complexes in reward processing and associative learning, processes known to rely heavily on these subregions in rodents.

In the first two studies, we provide evidence that hr-fMRI can be used to replicate findings from rodent lesion studies by showing different contributions of BLA and CeN in associative learning. First, as in rodents, we show that BLA contributes to specific Pavlovian Instrumental transfer (PIT), while CeN contributes to general PIT. We

then show that BLA is involved in encoding appetitive contexts while CeN is involved in encoding aversive ones, results which resonate with previous rodent studies.

In the next two studies, we show how hr-fMRI can also be used to go beyond animal studies and test for specific computational hypotheses in humans. More specifically, using an instrumental conditioning task, we provide evidence for a differential contribution of BLA and CeN in the encoding of computational signals in reward and avoidance learning respectively. Finally, we use the precision of hr-fMRI to reveal that amygdala activity is more consistent with model-based learning (which incorporates a rich structure of the environment) than model-free learning (which does not incorporate such structure) regarding its underlying computations in Pavlovian conditioning, adding an extra layer of complexity to amygdala function.

This thesis highlights three main findings. First, BLA and CeN seem to make different contributions to specific versus general processes in associative learning. Second, BLA and CeN also seem to have a differential involvement in appetitive versus aversive processes. Finally, amygdalar subregions represent various computational signals required for learning and evidence points to the fact that their activity is better accounted for by a model-based approach. Overall, these findings suggest that CeN underlies simple conditioned motivational influences on behaviour, while BLA provides a more complex representational role in emotionally charged decisions and voluntary behaviour. Taken together, these works provide a major step forward in the circuit level understanding of this key part of the human brain but also reveal that given its complexity, a full picture of human amygdala function will only emerge with the continued use of hr-fMRI, which affords unique leverage on the functional properties of amygdalar subregions and promises to bridge the gap between the human and animal amygdala literatures.

PUBLICATIONS

The following is a list of manuscripts and abstracts arising from the experiments carried out throughout the duration of this Ph.D that have been or will be submitted for publication.

Thesis chapters are derived from the following manuscripts:

Chapter 3

Prévost, C., Liljeholm, M., Tyzka, M., & O'Doherty, J.P. (accepted at *Journal of Neuroscience*) Neural correlates of specific and general Pavlovian to instrumental transfer within human amygdalar subregions: a high-resolution fMRI study.

Chapter 4

Prévost, C., McNamee D.C., O'Doherty, J.P. (in prep.) Differential involvement of amygdalar subregions in context-learning: evidence from high-resolution imaging.

Chapter 5

Prévost, C., McCabe, J.A., Jessup, R.K., Bossaerts, P., & O'Doherty, J.P. (2011) Differentiable contributions of human amygdalar subregions in reward and avoidance learning. *European Journal of Neuroscience*, 34(1):134-45, selected as a "Featured article"

Chapter 6

Prévost, C., Jessup, R.K., McNamee D.C., Bossaerts, P., & O'Doherty, J.P. (under review) Evidence for Model-based Computations in the Human Amygdala during Pavlovian Conditioning.

Published abstracts:

Prévost, C., McCabe, J.A., Jessup, R.K., Bossaerts, P., & O'Doherty, J.P. (2010, poster) Differentiable contributions of human amygdalar subregions in reward and avoidance learning. Society for Neuroscience, San Diego, California.

Prévost, C., McCabe, J.A., Jessup, R.K., Bossaerts, P., & O'Doherty, J.P. (2010, poster) Differentiable contributions of human amygdalar subregions in reward and avoidance learning. Human Brain Mapping, Barcelona, Spain.

Prévost, C., McCabe, J.A., Jessup, R.K., Bossaerts, P., & O'Doherty, J.P. (2010, poster) Differentiable contributions of human amygdalar subregions in reward and avoidance learning. Motivational & Cognitive Control, Oxford, England.

Review:

Prévost, C., & O'Doherty, J.P. (in prep.) New insights into the human amygdala circuitry of associative learning using high-resolution imaging.

TABLE OF CONTENTS

	PAGE
DECLARATION	i
ACKNOWLEDGEMENTS	ii
SUMMARY	iv
PUBLICATIONS	vi
TABLE OF CONTENTS	vii
ABBREVIATIONS	xiv
LIST OF TABLES	xv
LIST OF FIGURES	xvi

CHAPTER 1

GENERAL INTRODUCTION	1
1.1 A bit of history about the amygdala	3
1.2 The amygdala: a structurally and functionally heterogeneous brain area	5
1.3 Neuroanatomy and connectivity of the amygdala	7
1.3.1 Neuroanatomy and connectivity of the rodent amygdala	8
1.3.1.1 BLA complex	10
1.3.1.2 CeN complex	13
1.3.2 Neuroanatomy and connectivity of the non-human primate amygdala	17
1.3.2.1 BLA complex	19
1.3.2.2 CeN complex	20
1.3.3 Neuroanatomy and connectivity of the human amygdala	20
1.3.3.1 Location of BLA and CeN	20
1.3.3.2 Connectivity of the human amygdala	22

1.4	Theories of amygdala function	24
1.4.1	The 'serial' view	24
1.4.2	The 'parallel' view	25
1.5	Theories of associative learning	26

CHAPTER 2

METHODS USED IN THE THESIS	31
2.1	Functional Magnetic Resonance Imaging (fMRI) 31
2.1.1	Physics of fMRI 32
2.1.2	Physiology of fMRI 34
2.1.3	Block, event-related and mixed designs 36
2.1.4	Issues associated with fMRI 38
2.2	High-resolution imaging of the human amygdala 40
2.2.1	The development of high-resolution imaging 41
2.2.2	The amygdala: a challenging structure to image 43
2.2.3	A recipe for high-resolution imaging of the amygdala 44
2.2.3.1	Decrease your voxel size 45
2.2.3.2	Unwarp your data 46
2.2.3.3	Specifically normalize the amygdalae 47
2.2.3.4	Do not smooth your data 48
2.2.3.5	Remove physiological fluctuations from your data 49
2.2.3.6	Outline your ROIs 49

CHAPTER 3

NEURAL CORRELATES OF SPECIFIC AND GENERAL PAVLOVIAN TO INSTRUMENTAL TRANSFER WITHIN HUMAN AMYGDALAR SUBREGIONS: A HIGH-RESOLUTION fMRI STUDY	52
3.1	Introduction 52
3.2	Material and Methods 55

3.2.1 Participants	55
3.2.2. Stimuli	55
3.2.3 Training	56
3.2.3.1 Instrumental training	56
3.2.3.2 Pavlovian training	58
3.2.4 Task description	59
3.2.4.1 Instrumental phase	59
3.2.4.2 Pavlovian phase	59
3.2.4.3 PIT phase	59
3.2.5 Behavioural measures	63
3.2.5.1 Physiological fluctuations	63
3.2.5.2 Response rates	64
3.2.5.3 Affective evaluations of food pictures	64
3.2.6 Data Acquisition	64
3.2.7 Preprocessing	65
3.2.7.1 Amygdalae segmentation	66
3.2.7.2 Amygdalae normalization	66
3.2.8 fMRI data analysis	68
3.2.9 Plotting of parameter estimates	70
3.3 Results	71
3.3.1 Behavioural results	71
3.3.1.1. Subjective ratings of food rewards	71
3.3.1.2 Instrumental conditioning	71
3.3.1.3 Pavlovian conditioning	72
3.3.1.4 Pavlovian to Instrumental transfer effects	73
3.3.2 fMRI results	75
3.4 Discussion	78

CHAPTER 4

DIFFERENTIAL INVOLVEMENT OF AMYGDALAR SUBREGIONS IN CONTEXT-LEARNING: EVIDENCE FROM HIGH-RESOLUTION IMAGING

83

4.1 Introduction	83
4.2 Material and Methods	85
4.2.1 Subjects	85
4.2.2 Task description	86
4.2.3 Apparatus	89
4.2.4 Fluctuations in heart rate	89
4.2.5 Data Acquisition	89
4.2.6 Preprocessing	90
4.2.7 fMRI data analysis	90
4.2.8 ROI analyses	92
4.3 Results	92
4.3.1 Behavioural evidence of Pavlovian conditioning	92
4.3.2 fMRI results	92
4.4 Discussion	95

CHAPTER 5

DIFFERENTIABLE CONTRIBUTIONS OF HUMAN AMYGDALAR SUBREGIONS IN THE COMPUTATIONS UNDERLYING REWARD AND AVOIDANCE LEARNING

98

5.1 Introduction	98
5.2 Methods	100
5.2.1 Subjects	100
5.2.2 Task description	101
5.2.3 Prescan training	103
5.2.4 Data acquisition	104

5.2.5 Preprocessing	105
5.2.6 Computational model-based analysis	107
5.2.6.1 Action value learning	107
5.2.6.2 Action selection computation	109
5.2.6.3 State value learning	110
5.2.7 fMRI data analysis	112
5.2.8 ROI analyses	114
5.3 Results	115
5.3.1 Behavioural results	115
5.3.2 fMRI results	117
5.3.2.1 Action value signals	118
5.3.2.2 Reward prediction error signals	119
5.3.2.3 State value signals	121
5.3.2.4 Estimation uncertainty signals	121
5.4 Discussion	124
5.5 Conclusion	130

CHAPTER 6

EVIDENCE FOR MODEL-BASED COMPUTATIONS IN THE HUMAN AMYGDALA DURING PAVLOVIAN CONDITIONING	131
6.1 Introduction	131
6.2 Experimental Procedures	135
6.2.1 Subjects	135
6.2.2 Task description	135
6.2.3 Subject instructions	139
6.2.4 Apparatus	140
6.2.5 Behavioural measures	140
6.2.5.1 Affective evaluations of the fractal images and liquids	140

6.2.5.2 Preference ranking test	141
6.2.5.3 Pupillary dilation	141
6.2.5.4 Fluctuations in respiration and heart rate	142
6.2.6 Data Acquisition	142
6.2.7 Preprocessing	143
6.2.8 Computational model analysis	143
6.2.8.1 Model-based learning algorithms	144
6.2.8.2 Model-free learning algorithms	147
6.2.8.3 Model comparison on behavioural data	149
6.2.9 fMRI data analysis	152
6.2.10 Model comparison on BOLD data	153
6.2.11 ROI analyses	154
6.3 Results	155
6.3.1 Behavioural results	155
6.3.1.1 Affective ratings for the liquid outcomes	155
6.3.1.2 Revealed preference rankings for the cue stimuli	155
6.3.1.3 Pleasantness ratings for the cue stimuli	156
6.3.1.4 Pupil diameter and blink rate changes as a function of conditioning	157
6.3.1.5 Changes in pulse and respiration patterns as a function of conditioning	157
6.3.1.6 Model Comparison on Behavioural Data	158
6.3.2 fMRI results	160
6.3.2.1 Model-based learning algorithm: HMM model	160
6.3.2.2 Model-free learning algorithms	163
6.3.2.3 Model-comparison on BOLD data	164
6.4 Discussion	165

CHAPTER 7	
GENERAL DISCUSSION	171
7.1 Elucidating the functions of human amygdalar subregions	171
7.1.1 Summary of the thesis findings	172
7.1.2 Specific versus general processes in the amygdala	172
7.1.3 Differential contribution of BLA and CeN in appetitive versus aversive representations	175
7.1.4 The amygdala, a structure implementing model-based learning?	177
7.1.5 Reconciling our findings with the animal literature	179
7.2 Limitations of hr-fMRI	182
7.2.1 Limitations of our hr-fMRI protocol	182
7.2.1.1 Defining the amygdala	182
7.2.1.2 Segmentation of amygdalar subregions	183
7.2.1.3 Coupling model-based fMRI with hr-fMRI of the amygdala	186
7.2.2 The amygdala: a polynuclear structure	187
7.2.3 Limitations inherent to hr-fMRI (and fMRI)	188
7.3 Challenges and future directions for hr-fMRI	190
7.3.1 Combining hr-fMRI with other approaches	190
7.3.2 Investigating the interaction between amygdalar subregions and other brain structures	192
7.3.3 Application to the human clinical domain	194
7.4 Concluding comments	196
REFERENCES	198

ABBREVIATIONS USED IN THE THESIS

ANOVA Analysis of Variance
BIC Bayesian Information Criterion
BMS Bayesian Model Selection
BLA Basolateral complex
BOLD Blood Oxygen Level Dependent
CBV Cerebral Blood Volume
CeN Centromedial complex
Co Cortical nucleus
CR Conditioned Response
CS Conditioned Stimulus
DTI Diffusion Tensor Imaging
DWI Diffusion Weighted Imaging
EEG ElectroEncephaloGram
EPI Echo Planar Imaging
ER Expected Reward
EV Expected Value
ERP Event Related Potential
EU Estimation Uncertainty
fMRI functional Magnetic Resonance Imaging
GLM General Linear Model
HMM Hidden Markov Model
HRF Haemodynamic Response Function
Hr-fMRI High-resolution fMRI
ITI Inter-Trial Interval
LFP Local Field Potential
MEG MagnetoEncephaloGram
MNI Montreal Neurological Institute
MR Magnetic Resonance
MRI Magnetic Resonance Imaging
MTL Medial Temporal Lobe
NAc Nucleus Accumbens
NLL Negative Log Likelihood
PET Positron Emission Tomography
PIT Pavlovian-Instrumental Transfer
PH Pearce-Hall
PFC PreFrontal Cortex
rCBF regional Cerebral Blood Flow
RF RadioFrequency
RL Reinforcement Learning
ROI Region of Interest
ROI-AL ROI Alignment
RR Random Ratio
RT Reaction Time
RW Rescorla-Wagner
SNR Signal to Noise Ratio
SSE Sum Squared Error
SVC Small Volume Correction
TE Echo Time
TMS Transcranial Magnetic Stimulation
TR Repetition Time
VTA Ventral Tegmental Area

LIST OF TABLES

TABLE	PAGE
Table 3.1 Trial composition for the instrumental, Pavlovian and PIT phases	63
Table 3.2 Mean response rates (RR) $\pm$ SEM for each participant during the instrumental training and the PIT phase	75
Table 5.1 Model comparison between several variants of the delta learning model and a baseline model	117

LIST OF FIGURES

FIGURE	PAGE
Figure 1.1 Location of the amygdalae in the human brain	1
Figure 1.2 Brightfield photomicrographs of the rat amygdaloid complex	9
Figure 1.3 Summary of the cortical, subcortical, and intra-amygdaloid inputs and outputs from/to the lateral amygdala	11
Figure 1.4 Summary of the cortical, subcortical, and intra-amygdaloid inputs and outputs from/to the basolateral amygdala	12
Figure 1.5 Summary of the cortical, subcortical, and intra-amygdaloid inputs and outputs from/to the basomedial amygdala	13
Figure 1.6 Summary of the cortical, subcortical, and intra-amygdaloid inputs and outputs from/to the central amygdala	14
Figure 1.7 Summary of the cortical, subcortical, and intra-amygdaloid inputs and outputs from/to the medial amygdala	15
Figure 1.8 Schematic of the major inputs and outputs from/to the main amygdala nuclei	16
Figure 1.9 Nissl images of the primate amygdala nuclei	18
Figure 1.10 Schematic of the human amygdala nuclei	22
Figure 2.1 Magnetisation of water	32
Figure 2.2 The BOLD response	35
Figure 2.3 Schematic of task designs in fMRI	38
Figure 2.4 High-resolution anatomical images of human MTL	42
Figure 2.5 Illustration of our hr-fMRI protocol	51
Figure 3.1. PIT task illustration	57

Figure 3.2 Behavioural results	74
Figure 3.3 Blood oxygen level-dependent (BOLD) signal for the main effect of specific PIT	76
Figure 3.4 Specific and General PIT in the amygdala	78
Figure 4.1 Appetitive versus aversive Pavlovian learning task	88
Figure 4.2 Amygdala signals following new learning	94
Figure 5.1 Reward/avoidance instrumental learning task	103
Figure 5.2 Behavioural results	116
Figure 5.3 Action reward value signals	119
Figure 5.4 Reward prediction error signals	120
Figure 5.5 Estimation uncertainty signals	123
Figure 6.1 Appetitive versus aversive Pavlovian learning task	138
Figure 6.2 Computational models illustration	151
Figure 6.3 Behavioural results	159
Figure 6.4 Expected value signals from the model-based learning algorithm model in the amygdala	161
Figure 6.5 Precision signals from the model-based learning algorithm in the amygdala	162
Figure 7.1 Amygdalar subregions according to Amunts' probabilistic maps, our template and the Mai atlas	185

CHAPTER 1 INTRODUCTION

The amygdala, a brain region lying at the heart of the limbic system has the distinction of being the most densely interconnected structure of the primate forebrain (see Figure 1.1 for a general view of the human amygdala), with an impressive wide-ranging scope of influence.

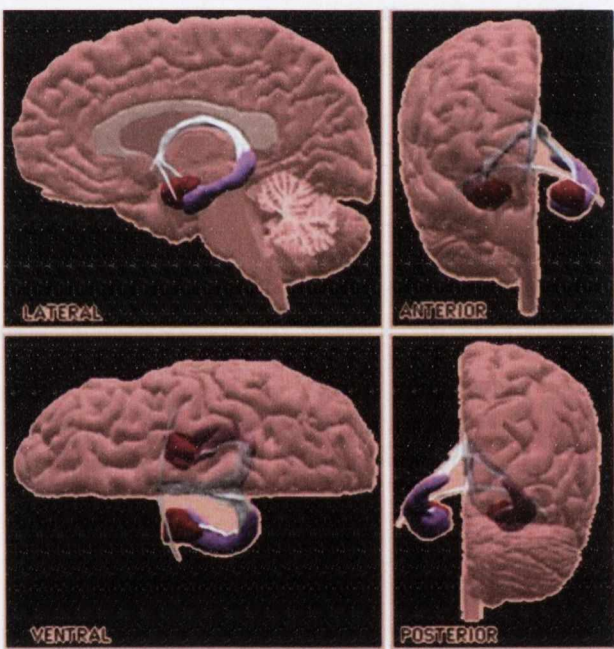


Figure 1.1. Location of the amygdalae (in red) in the human brain. Lateral (top left), anterior (top left), ventral (bottom left) and posterior (bottom right) views are shown.

Research on this small subcortical structure has blossomed in the last decade. During these years, new techniques and new questions have had an extraordinary impact on our understanding of the amygdala. The convergence of findings from both animal and human literatures has been a dominant theme that has been aided by

advances in brain imaging and the discovery of specific cognitive deficits due to amygdala damage in humans.

The numerous component nuclei of this inherently heterogeneous area receive highly specific projections from and send projections to an incredibly diverse number of brain regions suggesting some segregation of function across these components (see section 1.3 for detailed amygdala anatomy). This complex architecture thus precludes a simple attribution of function to the amygdala as a whole, an idea supported by rodent lesion studies that show different and doubly dissociable roles for the basolateral (BLA) and central (CeN) areas of the amygdala in associative learning (Corbit and Balleine, 2005; Holland and Gallagher, 1999, 2003; Killcross et al., 1997; Petrovich et al., 2009). However, the contribution of specific human amygdalar subregions to reward processing and associative learning (or indeed any other field) remains largely unknown.

Unlike animal research, human research relies heavily on state-of-the-art non-invasive techniques such as fMRI. However, the limited spatial resolution of fMRI has precluded reliable identification of the functions of human amygdalar subregions, making critical comparisons with the animal literature difficult. Despite this major limitation which barely allows for coarse spatial dissociations within the amygdala, a number of fMRI studies have attempted more fine-grained signal attribution between dorsal and ventral parts (Kim et al., 2004; Mackiewicz et al., 2006; Morris et al., 2001a; Whalen et al., 1998; Whalen et al., 2001), medial and lateral (Kim et al., 2003; Zald and Pardo, 2002), or anterior versus posterior parts of the amygdala (Gottfried et al., 2002; Morris et al., 2002). However, sequences used to image the amygdala using standard

fMRI are prone to problems of signal to noise ratio (SNR), inhomogeneities, anatomic variability and mislocalization (see section 1.6.2 for more details), rendering these signal attributions to various subareas of the amygdala suspect (as explained below, such claims require careful data analysis procedures which were not performed in these studies).

Therefore, increased anatomical precision is essential to further understanding of the functional heterogeneity that is undoubtedly present in the amygdaloid complex and will be crucial to guide future experimental questions and targeted hypotheses.

1.1 A BIT OF HISTORY ABOUT THE AMYGDALA

The term 'amygdala', referring to the subcortical grey matter found rostral to the hippocampus in the medial temporal lobe (MTL), is thought to have been used for the first time in 1819 by the anatomist K.F. Burdach. This name is believed to come from the almond (Greek meaning of 'amygdala') shape of the basolateral area of the amygdala. In 1923, the contemporary partition of the amygdala into its basolateral (BLA), centromedial (CeN) and cortical (Co) divisions was formally recognized by J.B. Johnstone. These divisions of the human amygdala, which also apply to other complex vertebrates, are justified on both functional and structural grounds (see Section 1.3).

However, in recent years, several researchers have attempted to rethink its organization and the anatomical boundaries of the amygdala have engendered new debate. Two major alternative views have been posited. The first one postulates that there is an extended amygdala, which refers to corridors of cells extending from the CeN through the sublenticular/basal forebrain, forming a continuum with the lateral and medial nuclei of the bed nucleus of the stria terminalis (Alheid and Heimer, 1988;

Heimer, 2003; Heimer et al., 1997). The second proposal argues that the nuclei of the BLA complex are nuclear extensions of the cortex while the CeN and cortical nucleus are ventral extensions of the striatum (Swanson and Petrovich, 1998). In this view, the amygdala consists of areas that belong to other systems of the brain, and as such it is neither a structural nor a functional unit.

The amygdala has an interesting pattern indicative of both mosaic and coordinated evolution (Stephan et al., 1987). The former implies that each area evolved separately with its own specific efferents and afferents whereas the latter indicates that the different areas evolved together. While the BLA complex and the neocortex show the same pattern of size differences across taxonomic groups (both being remarkably expanded in primates) indicative of coordinated evolution, the CeN complex displays no such taxonomic differences, indicating a degree of mosaic evolution of the amygdala. Furthermore, studies of correlated evolution have revealed that amygdaloid entities correlated most strongly with each other rather than with other brain areas, highlighting the functional and connectional integrity of this structure (Barton and Aggleton, 2000).

Originally, the structure and functions of the amygdala were thought to be primitive compared with the highly elaborated neocortex because it is an evolutionary ancient brain system. This simplistic conception of brain evolution has been challenged, and it is now clear that as brains evolved, new functions were built into and integrated with existing systems (Damasio, 1994). Emotional and cognitive systems in which the amygdala plays a major role have evolved together and are aspects of the unified process underlying adaptive behaviour. As Pinker once elegantly expressed: "Emotion is not just running away from a bear. It can be set off by the most sophisticated information processing the mind is capable of... And the emotions help to connive intricate plots for escape, revenge, ambition, and courtship" (Pinker, 1998).

1.2 THE AMYGDALA: A STRUCTURALLY AND FUNCTIONALLY HETEROGENEOUS BRAIN AREA

The animal literature has pointed out that, rather than being a single functionally or anatomically integrated structure, the amygdaloid complex is a heterogeneous conglomeration of nuclear groups (Swanson and Petrovich, 1998). Each of these anatomically distinct nuclei has its own characteristic functions and pattern of connectivity (see Section 1.3). However, the amygdala is commonly divided into a 'cortical' complex which includes the lateral, basolateral and basomedial nuclei (BLA), and a 'striatal' complex including the medial and central nuclei (CeN).

The amygdala is an integrative neurobiological structure subserving wide-ranging adaptive functions that cross the categorical boundaries of motivation, arousal, emotion and cognitive processes such as attention, learning and memory (Davis and Whalen, 2001; Gallagher and Holland, 1994; Zald, 2003). Recently, it has also received considerable attention in the social domain in humans (Adolphs, 2010). Grossly generalized statements such as the amygdala being 'the emotional center of the brain' despite amygdala lesions largely sparing important aspects of emotional processing have long been disseminated. These claims mostly originate in the animal literature, which has confined its research primarily to fear conditioning, a form of Pavlovian conditioning that involves pairing an initially neutral stimulus (conditioned stimulus or CS) with an aversive stimulus (unconditioned stimulus or US). Fear conditioning results in the exhibition of a fear conditioned response (CR) in the presence of the CS. As an extremely robust form of learning, this model system has lent itself well to neural analyses at the behavioural, cellular and molecular levels (Davis and Shi, 2000; Phelps and LeDoux, 2005; Rodrigues et al., 2004). This widely studied paradigm has also led to the erroneous idea that the amygdala is predominantly involved in processing

negative reinforcement to produce negative affect. An integrative approach emerged in the last decade, preventing amygdala function from being exclusively viewed through the lens of a single behavioural model. However, probably arising as an artifact of research interests or the significant ease with which negative stimuli induce strong motivational states, fear is still the emotion best understood in terms of brain mechanisms.

Although substantially smaller than the literature on aversive conditioning, a relatively large body of research has focused on the amygdala's involvement in appetitive conditioning and reward learning (and indeed this thesis concentrates on reward learning), particularly on how rewards are processed and used to motivate and reinforce behaviour in both animals and humans (Baxter and Murray, 2002; Cador et al., 1989; Cardinal et al., 2002; Everitt et al., 1999; Holland and Gallagher, 2004; Parkinson et al., 2000). Furthermore, electrophysiological, neuropharmacological and lesion studies in animals have provided strong evidence for the involvement of the amygdala in the evaluation and response to emotionally significant sensory signals. Similarly, in humans, fMRI and neuropsychological studies have shown a role for the amygdala in processing aversive and positively valenced stimuli from multiple sensory modalities (Birbaumer et al., 1998; Hamann et al., 2002; Irwin et al., 1996; Morris et al., 2001a; O'Doherty et al., 2001; Taylor et al., 2000; Yamasaki et al., 2002; Zald et al., 2002; Zald and Pardo, 2002), reflecting a common multimodal feature of amygdala coding. Simultaneously, neuropsychologists have developed sophisticated paradigms to investigate which aspects of emotional processing were affected by amygdala lesions. Taken together, these lines of research have laid the groundwork for refining theories of amygdala function.

However, due to limitations in the spatial resolution of current available tools to study the human amygdala, the term amygdala has been used to designate an anatomical area as opposed to a functionally integrated circuit. Despite the fact that this limitation has precluded progress towards a comprehensive understanding of amygdala function, the ability to map responses to this region as a whole has provided an immense source of information on the functions subserved by the human amygdala. However, there exists a strong need for improved spatial resolution and appropriate techniques in order to move toward a far more refined understanding of the amygdala's functional properties. This ambitious goal has been the motivation for the work presented in this thesis.

1.3 NEUROANATOMY AND CONNECTIVITY OF THE AMYGDALA

The amygdala is a multinuclear complex comprised of a large number of anatomically-distinct nuclei which differ cytoarchitectonically, chemoarchitectonically and connectionally. In total, thirteen nuclei and their numerous subdivisions compose the rodent and the primate amygdaloid complex (Amaral and Bassett, 1989; Amaral, 1992; Freese and Amaral, 2009; Pitkänen, 2000; Pitkänen and Amaral, 1998; Price, 1986). However, as mentioned previously, the animal literature has ordinarily congregated these nuclei into a basolateral area which includes the lateral, basolateral and basomedial nuclei (BLA), a centromedial area which includes the medial and central nuclei (CeN), and the cortical nucleus (Co). This congregation appears especially evident in rodent lesion studies where care has been taken to investigate BLA and CeN function separately (note that Co has not been systematically investigated). Taken together, this research has established that BLA and CeN fulfill

different functions, at least in rodents (Balleine and Killcross, 2006; Corbit and Balleine, 2005; Holland and Gallagher, 2003; Killcross et al., 1997; LeDoux, 2007; Petrovich et al., 2009; Rogan and LeDoux, 1996). On these grounds, particular emphasis will be given to the BLA and CeN area in this section as well as throughout this thesis.

Anterograde and retrograde studies have shown that each nucleus (and even their subdivisions) has unique interconnections with other amygdala nuclei and a wide-reaching network of projections with a diverse array of brain regions (Freese and Amaral, 2009; Pitkänen, 2000). Given the large number of amygdaloid nuclei and the fact that specific connectivity patterns define function, differences in connectivity among amygdala nuclei are crucial to the identification of functional entities, and knowledge of the connectivity of the rat and monkey amygdaloid complex is therefore critical to continued progress in understanding the functions of the human amygdala.

This section provides a summary of the principles of the rodent, non-human and human primate amygdala nuclei location and wiring. A thorough description of the intra-nuclear, intra-amygdaloid, inter-amygdaloid, extra-amygdaloid inputs and outputs from both the rodent and primate amygdalae is beyond the scope of this thesis; however, these connections are catalogued in a comprehensive manner elsewhere (Freese and Amaral, 2009; Pitkänen, 2000).

1.3.1 Neuroanatomy and connectivity of the rodent amygdala

In this section, we focus on the location and key connections of the main amygdala nuclei within the BLA and CeN complexes. Different coronal views of the rat amygdaloid complex are shown in Figure 1.2.

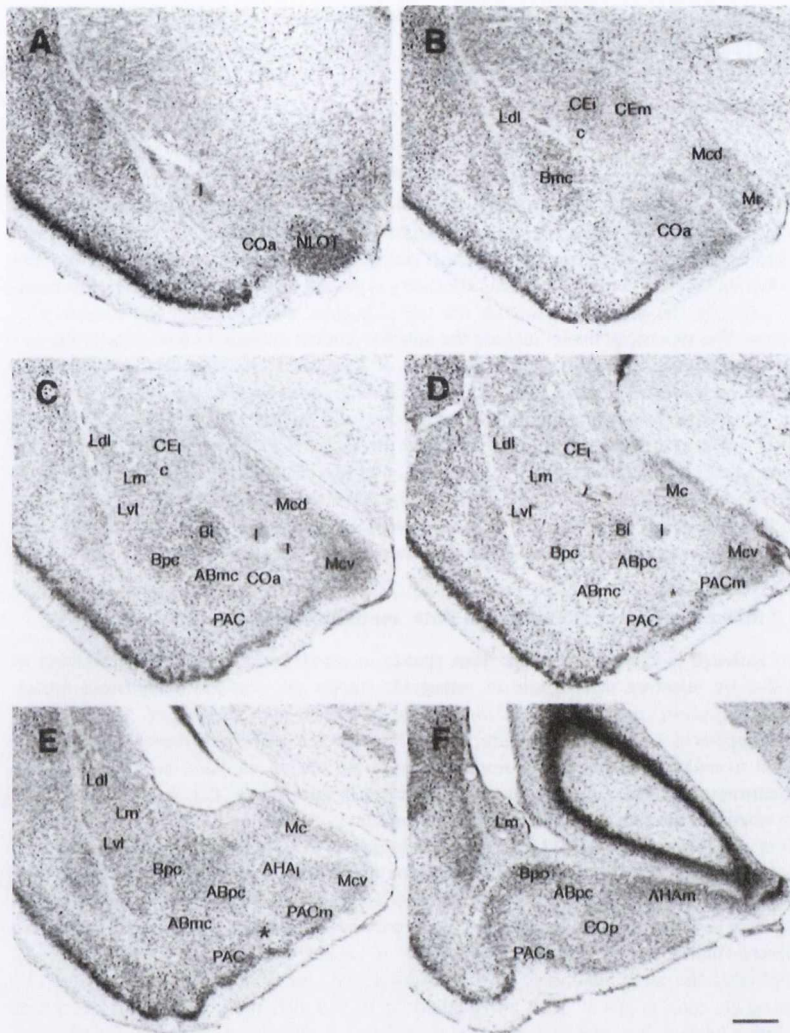


Figure 1.2. Brightfield photomicrographs from thionin-stained coronal sections of the rat amygdaloid complex showing the location of various amygdaloid nuclei and nuclear divisions. Six rostrocaudal slices are presented (A is the most rostral and F the most caudal). (I) intercalated nuclei, (COa) anterior cortical nucleus, (NLOT) nucleus of the lateral olfactory tract, (Ldl) dorsolateral division of the lateral nucleus, (CEi) intermediate division of the central nucleus, (CEm) medial division of the central nucleus, (Mcd) dorsal part of the medial nucleus, (Mr) rostral division of the medial nucleus, (Bmc) magnocellular division of the basolateral nucleus, (Mcv) ventral part of the medial nucleus, (PAC) periamygdaloid cortex, (ABmc) magnocellular division of the basomedial nucleus, (Bpc) parvicellular division of the basolateral nucleus, (Bi)

intermediate division of the basolateral nucleus, (Lvl) ventrolateral division of the lateral nucleus, (Lm) medial division of the lateral nucleus, (ABpc) parvicellular division of the basomedial nucleus, (Mc) caudal division of the medial nucleus, (AHAl) lateral division of the amygdalo-hippocampal area, (Cop) posterior cortical nucleus, (AHAm) medial division of the amygdalo-hippocampal area. Adapted from Pitkanen (2000).

1.3.1.1 BLA complex

- Lateral nucleus: dorsally located in the rat amygdala, the lateral nucleus is lateral to the central nucleus and the lateral ventricle, medial to the external capsule and dorsal to the basal nucleus (Figure 1.2). The main feature of the lateral nucleus is that it receives heavy inputs from a larger number of sensory-related lateral cortical areas (including the gustatory, olfactory, visceral, somatosensory, auditory and visual cortical areas) than any other amygdala nucleus (McDonald, 1998; Shi and Cassell, 1997, 1998a, b, 1999) and as such, it is generally viewed as the sensory interface of the amygdala (Amaral, 1992; Heimer and Van Hoesen, 2006; LeDoux et al., 1990; McDonald, 1998; Turner and Herkenham, 1991). Although it also receives projections from the prefrontal cortex (PFC), these are markedly lighter than the projections to the basolateral nucleus. Furthermore, the lateral nucleus receives extensive inputs from the other amygdala nuclei, hippocampal formation, midline and posterior thalamic nuclei and hypothalamus. Conversely, it provides considerable projections to the other amygdala nuclei, hippocampal formation and perirhinal cortex, and PFC. A summary of these connections is presented in figure 1.3.

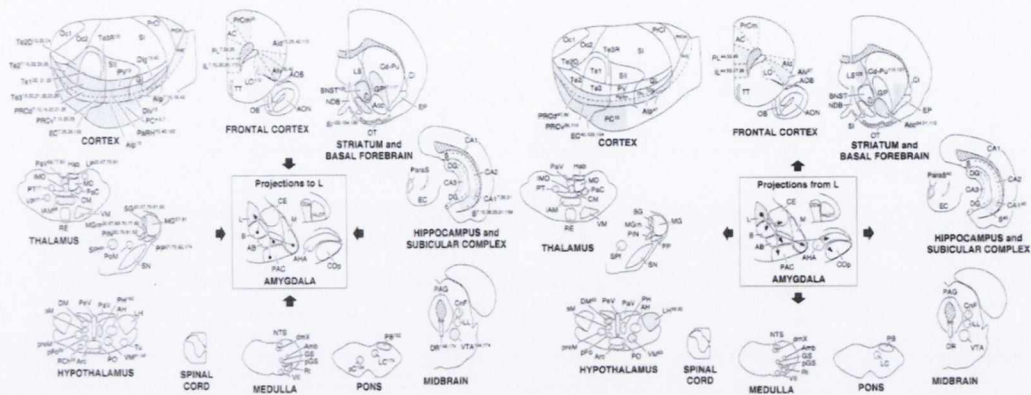


Figure 1.3. Summary of the cortical, subcortical, and intra-amygdaloid inputs (left) and outputs (right) from/to the lateral nucleus. Only the areas shaded are providing the inputs (left) or outputs (right). (L) lateral nucleus, (B) basolateral nucleus, (AB) basomedial nucleus, (CE) central nucleus, (M) medial nucleus. From Pitkänen (2000).

- Basolateral nucleus: in its rostral part, the basal nucleus is located rostral to the lateral nucleus, medial to the external capsule and lateral to the central nucleus. More caudally, it is located ventral to the lateral nucleus and dorsal to the basomedial nucleus, and it is lateral to the lateral nucleus in its most caudal part (Figure 1.2). The basolateral nucleus receives heavy inputs from the medial PFC. In return, projections from the basolateral nucleus to the medial and lateral PFC are the most widespread among amygdala nuclei. In addition, it is the main source of amygdala inputs to the nucleus accumbens, caudate nucleus and putamen, striatal areas involved in the control of instrumental behaviours (Everitt et al., 1989; Fudge et al., 2002; Pitkänen et al., 1997; Russchen et al., 1985). It is also the most important target of the hippocampal formation to the amygdala, and, likewise, the basolateral nucleus provides the most widespread amygdaloid projections to the hippocampal formation. Finally, the basolateral nucleus also receives inputs from sensory-related cortical areas and

projects to the bed nucleus of the stria terminalis and the substantia innominata. A summary of these connections is presented in figure 1.4.

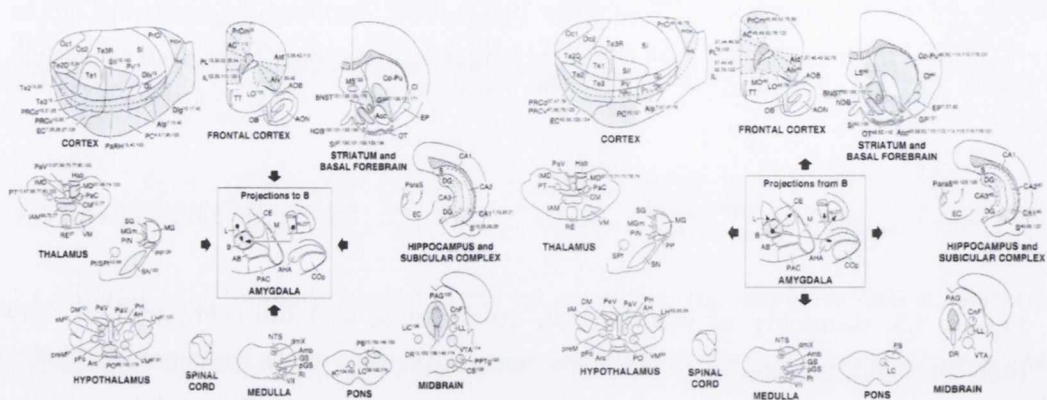


Figure 1.4. Summary of the cortical, subcortical, and intra-amygdaloid inputs (left) and outputs (right) from/to the basolateral nucleus. Only the areas shaded are providing the inputs (left) or outputs (right). (L) lateral nucleus, (B) basolateral nucleus, (AB) basomedial nucleus, (CE) central nucleus, (M) medial nucleus. From Pitkänen (2000).

- Basomedial nucleus: bordered by the basolateral nucleus dorsally and the periamygdaloid cortex ventrally, the basomedial nucleus is located dorsal to the amygdalohippocampal area in its caudal part (Figure 1.2). It receives its main projections from the lateral sensory-related cortical areas, hippocampal formation, medial PFC and some thalamic and hypothalamic nuclei. Its heaviest outputs terminate in other amygdala nuclei, sensory-related cortical areas, medial PFC, hippocampal formation, hypothalamus, bed nucleus of stria terminalis, substantia innominata, nucleus accumbens, caudate nucleus and putamen. A summary of these connections is presented in figure 1.5.

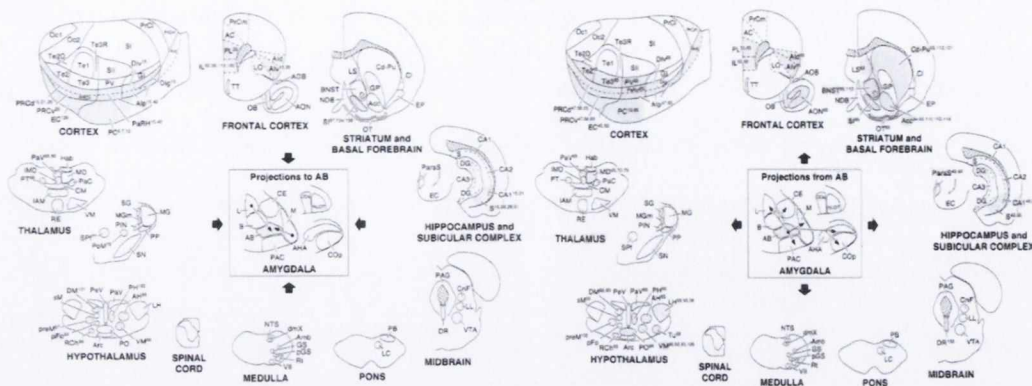


Figure 1.5. Summary of the cortical, subcortical, and intra-amygdaloid inputs (left) and outputs (right) from/to the basomedial nucleus. Only the areas shaded are providing the inputs (left) or outputs (right). (L) lateral nucleus, (B) basolateral nucleus, (AB) basomedial nucleus, (CE) central nucleus, (M) medial nucleus. From Pitkänen (2000).

1.3.1.2 CeN complex

- Central nucleus: located in the dorsomedial part of the rostral half of the amygdala, the central nucleus is medial to the lateral and basomedial nuclei and lateral to the stria terminalis. In its caudal part, it ends when the lateral ventricle appears (Figure 1.2). The central nucleus receives most of the projections from the midbrain, pons and medulla to the amygdala and provides in return most of the amygdala projections to these areas, thought to be involved in controlling emotional reactions (it receives the richest dopaminergic innervation among amygdala nuclei). It also receives and provides heavy inputs from/to numerous hypothalamic nuclei. In addition, this nucleus receives substantial inputs from the lateral sensory-related cortical areas, hippocampal formation, medial and lateral PFC, bed nucleus of stria terminalis, substantia innominata and some thalamic nuclei, and provides moderate to heavy inputs to the bed nucleus of stria terminalis in return. This nucleus is believed to act primarily as an

output station from the amygdala to other brain structures, and engender appropriate behavioural responses to information entering the amygdala. It has very few direct projections to the cerebral cortex; however, it can have a profound indirect action on cortical function via its projections to major neuromodulatory systems such as the ventral tegmentum area (dopamine), the raphe nuclei (serotonin), the nucleus basalis of Meynert (acetylcholine) and the locus coeruleus (norepinephrine). A summary of these connections is presented in figure 1.6.

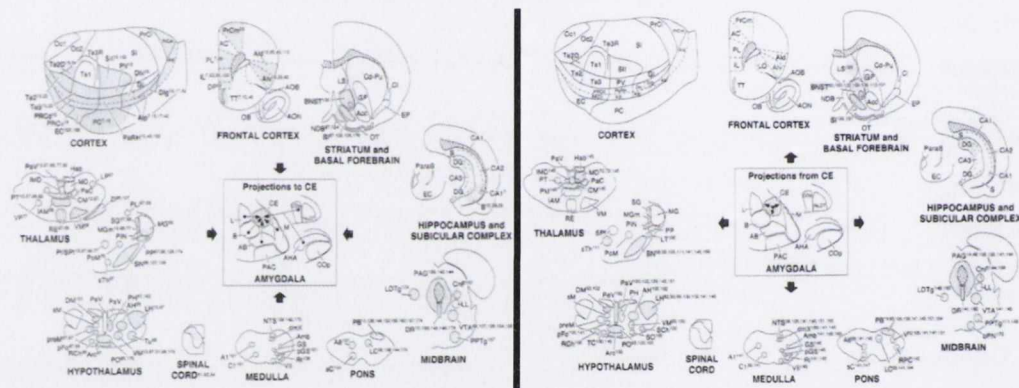


Figure 1.6. Summary of the cortical, subcortical, and intra-amygdaloid inputs (left) and outputs (right) from/to the central nucleus. Only the areas shaded are providing the inputs (left) or outputs (right). (L) lateral nucleus, (B) basolateral nucleus, (AB) basomedial nucleus, (CE) central nucleus, (M) medial nucleus. From Pitkänen (2000).

- Medial nucleus: beginning at the level of the nucleus of the lateral olfactory tract, the medial nucleus extends caudally until the lateral ventricle appears (Figure 1.2). Most of its projections originate in the PFC, bed nucleus of the stria terminalis and hypothalamus. Conversely, its major outputs terminate in other amygdaloid nuclei, the

olfactory system, the bed nucleus of the stria terminalis, the thalamus and hypothalamus. A summary of these connections is presented in figure 1.7.

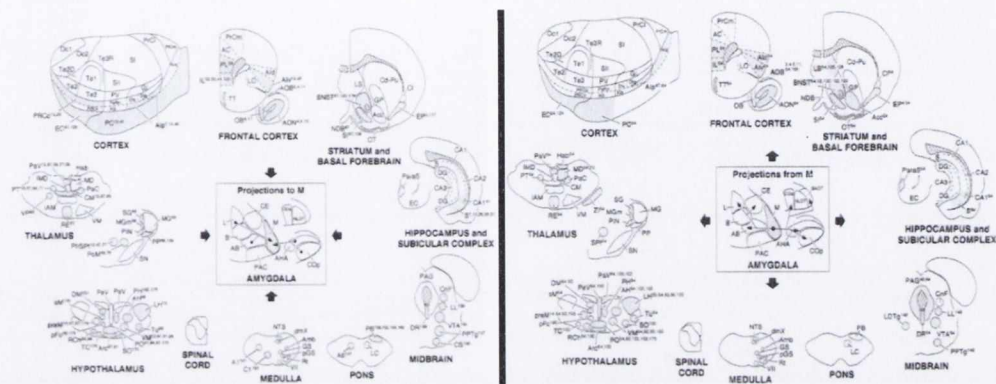


Figure 1.7. Summary of the cortical, subcortical, and intra-amygdaloid inputs (left) and outputs (right) from/to the medial nucleus. Only the areas shaded are providing the inputs (left) or outputs (right). (L) lateral nucleus, (B) basolateral nucleus, (AB) basomedial nucleus, (CE) central nucleus, (M) medial nucleus. From Pitkänen (2000).

A global summary of all these main connections is presented in figure 1.8. As shown, each amygdaloid nucleus is connected to a unique set of other brain regions.

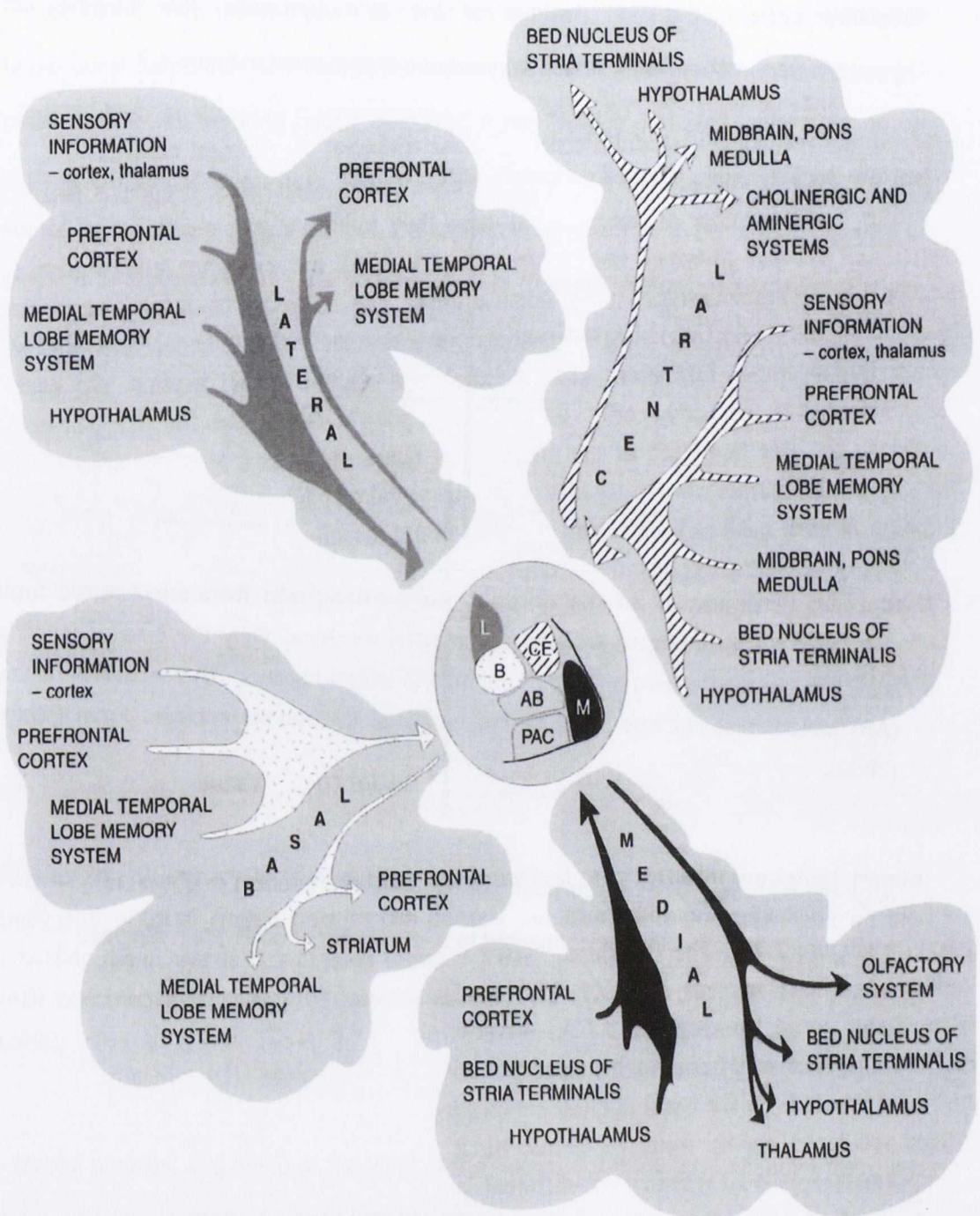


Figure 1.8. Schematic of the major inputs and outputs from/to the main amygdala nuclei. (L) lateral nucleus, (B) basolateral nucleus, (AB) basomedial nucleus, (CE) central nucleus, (M) medial nucleus. From Pitkänen (2000).

1.3.2 Neuroanatomy and connectivity of the non-human primate amygdala

In primates, the amygdala is rotated in a clockwise direction as compared to the rat amygdala, so that the lateral nucleus has both a dorsal and a ventral component. Therefore, the CeN lies dorsal to the BLA, creating a dorsal ventral distinction which is specific to the primate (human and non-human) amygdala (Heimer and Van Hoesen, 2006).

The primate amygdala has similar patterns of connectivity to that of the rat (Amaral, 1992). However, there exists some differences. The main one results from evolutionary reorganization, whereby the lateral nucleus is proportionally much larger in the non-human primate and human amygdala than in that of the rodent amygdala (Barger et al., 2007). This reorganization can be easily explained by the fact that the neocortex, which provides the majority of its projections to the lateral nucleus, has undergone the most elaboration in the primate brain (McDonald, 1998; Stephan et al., 1987).

To avoid redundancy with the previous section on rat amygdala anatomy, only the main characteristics of the primate amygdala nuclei within the BLA and CeN complex and the connections that may differ from that of the rodent amygdala are described. The location of the various primate nuclei are presented in figure 1.9.

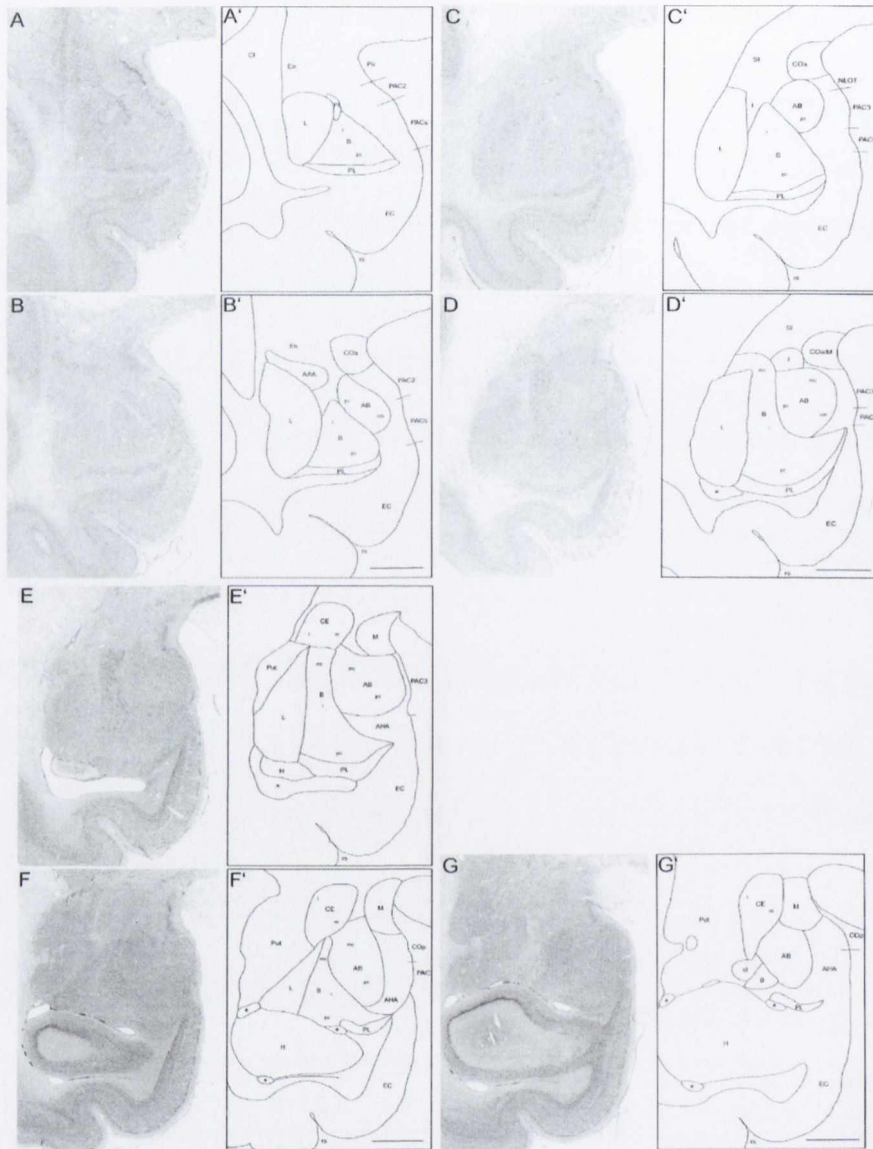


Figure 1.9. Nissl images of the primate amygdala nuclei from rostral (A) to caudal (G), sectioned in the coronal plane. Line drawings (A'-G') representing the borders of nuclei and subdivisions accompany each Nissl section. (AAA) anterior amygdaloid area, (AHA) amygdalo-hippocampal area, (L) lateral nucleus, (B) basolateral nucleus, (AB) basomedial nucleus, (CE) central nucleus, (M) medial nucleus, (CO) cortical nucleus, (EC) entorhinal cortex, (H) hippocampus, (I) intercalated nucleus, (NLOT) nucleus of the lateral olfactory tract, (PAC) periamygdaloid cortex, (Pir) piriform cortex, (PL)

paralamina nucleus, (Put) putamen, (SI) substantia innominate, (st) stria terminalis. From Freese and Amaral (2009).

1.3.2.1 BLA complex

- Lateral nucleus: the main characteristic of this nucleus is its strong reciprocal connections with visual, auditory and multimodal sensory areas. It also has reciprocal connections with the insular cortex and cingulate cortex and projects to the thalamus and entorhinal cortex.
- Basolateral nucleus: the key reciprocal connections of this nucleus are with the medial PFC and the orbitofrontal cortex, although there are more efferents from the amygdala to the PFC than from the PFC to the amygdala. Another important characteristic is that it originates most of the projections to the striatum, with its parvocellular division projecting to the medial part of the nucleus accumbens, and its magnocellular division projecting to the tail and body of the caudate nucleus and putamen (Russchen et al., 1985). It also has very strong reciprocal connections with the insular cortex, the cingulate cortex and the hippocampal formation. Finally, it receives heavy projections from the basal nucleus of Meynert and projects to the bed nucleus of the stria terminalis, the thalamus, the occipital cortex and the basal forebrain.
- Basomedial nucleus: its main projections are sent to the insular cortex, the bed nucleus of the stria terminalis, the thalamus and the basal forebrain. In addition, it has important reciprocal connections with the hippocampal formation.

1.3.2.2 CeN complex

- Central nucleus: its main characteristic is its heavy reciprocal connections with the hypothalamus, the ventral tegmental area and the substantia nigra (from which it receives a strong dopaminergic input). Furthermore, it is the only amygdala nucleus with substantial reciprocal connections with the pons and medulla, thought to facilitate its role in autonomic response. It also receives projections from the insular cortex and projects to the bed nucleus of the stria terminalis and the basal forebrain.
- Medial nucleus: most notably, this nucleus projects to the medial part of the bed nucleus of the stria terminalis.

1.3.3 Neuroanatomy and connectivity of the human amygdala

In this section, the location of the human amygdala on the one hand and that of the BLA and CeN complexes on the other hand are first described. Issues regarding connectivity of the human amygdala are discussed next.

1.3.3.1 Location of BLA and CeN

The cytoarchitectonic organization of the human and macaque monkey amygdala is virtually completely homologous between the two species, suggesting that the primate amygdala is a reasonable proxy for the human amygdala (Pitkänen and Kemppainen, 2002; Sorvari et al., 1995; Sorvari et al., 1996a, b). In its rostral part, the human amygdala is bordered ventromedially by the entorhinal cortex, ventrally by the temporal horn of the lateral ventricle and subamygdaloid white matter and laterally by

white matter of the temporal lobe. Mid-rostrocaudally, the human amygdala increases in size and is bordered ventromedially by a thin tract of white matter separating the amygdala and the entorhinal cortex, laterally by the white matter of the temporal lobe and medially by the semiannular sulcus. Caudally, the human amygdala is bordered dorsally by the substantia innominata and fibers of the anterior commissure, laterally by the putamen, ventrally by the temporal horn of the lateral ventricle and the alveus of the hippocampus and medially by the optic tract (Mai et al., 2008).

In its most rostral part, the human amygdala is exclusively composed of the basolateral complex. The cortical nucleus appears in the dorso-medial part of mid-rostral amygdala. The centromedial complex appears slightly more caudally than the cortical nucleus in the most dorsal part of the amygdala. The basolateral complex increases in size as one moves caudally from the anterior amygdala, has its maximal size midrostrocaudally and then decreases as one moves further back toward the caudal amygdala, whereas the cortical nucleus and centromedial complex slightly enlarge midrostrocaudally, but do not decrease in size as one moves further caudally within the amygdala. The cortical nucleus ends midcaudally, the basolateral complex ends in caudal amygdala while the centromedial complex ends in the most caudal part of amygdala (Mai et al., 2008). Detailed schematics showing the location of the different amygdala nuclei are presented in figure 1.10.



Figure 1.10. Schematic of the human amygdala nuclei from rostral (top left) to caudal (bottom right), sectioned in the coronal plane. (L) lateral nucleus, (BL) basolateral nucleus, (BM) basomedial nucleus, (C) central nucleus, (M) medial nucleus, (Co) cortical nucleus. Adapted from Mai et al. (2008).

1.3.3.2 Connectivity of the human amygdala

Because of technical reasons, the connectivity of the human amygdala is largely unknown. Therefore, the non-human primate amygdala connectivity analyses provide the most accurate approximation for the human amygdala to date.

Nonetheless, although a fully coherent picture of human amygdala connectivity remains lacking, some efforts have recently been made in order to unravel connections between the amygdala as a whole and other brain structures. Using fiber tracking based on diffusion-weighted imaging (DWI) in combination with functional MRI, it was found that white matter tracts from the amygdala to the hippocampus, orbitofrontal cortex, and ventral striatum predicted how subjects adapted their behaviour following

positive and negative feedback (Cohen et al., 2008b). Probabilistic fiber tracking has also been used in combination with diffusion tensor imaging (DTI) to reveal structural connectivity between the amygdala and the medial and lateral orbitofrontal cortices, the anterior cingulate cortex and the dorsolateral prefrontal cortex (Bracht et al., 2009).

However, as the organization of information outflow from the amygdala emerging from anatomy research predicts important features of the functional organization of the different nuclei amygdaloid complex, it becomes clear that the specific contribution of these nuclei is difficult to ignore when trying to understand different aspects of amygdala-mediated behaviour. Therefore, although these first attempts looking at the whole amygdala are promising, this area of research remains ripe for further exploration.

In summary, the organization of afferent, efferent and intra-amygdaloid connections for each of the amygdaloid nuclei is characterized by a very fine-tuned topography and it appears evident that the amygdala is at the interface of important information exchange between a large number of functional systems of the brain. Outputs from different amygdala nuclei terminate mostly in non-overlapping areas, suggesting a high level of specialization for each of these nuclei. Overall, the organization of amygdala projections suggests that amygdala-evoked cortical (most notably sensory-related and PFC) and striatal modulation is largely generated by the BLA complex (i.e. the lateral, basolateral and basomedial nuclei) whereas endocrine and autonomic modulation occurs via the CeN complex (i.e. the central and medial nuclei).

The human literature leaves a host of connections unexplored; there are very few connectivity analyses in humans. Structural and functional MRI have only started to provide information about anatomical connectivity and estimates of functional connectivity between the human amygdala and other brain structures. Although a substantial number of functional aspects of the amygdala as a whole have been investigated, the complex architectural organization of the amygdala makes evident that functional studies of the human amygdala would greatly benefit from analyses at a finer level. Therefore, there is an important need for the separate roles of specific human amygdalar subregions to be distinguished.

1.4 THEORIES OF AMYGDALA FUNCTION

1.4.1 The 'serial' view

Animal models have offered a theoretical framework for furthering our understanding of amygdala function in human behaviour. Current theories of amygdala function have highlighted two major conflicting views. The early and probably most influential one arose from studies of Pavlovian fear conditioning in rodents, a form of aversive learning (LeDoux, 2000a; Maren, 2005). In this model, associations between sensory-perceptual stimuli and biologically significant stimuli are formed in the lateral amygdala (within the BLA), triggering the performance of numerous defensive reflexes via output through the central amygdala (within the CeN) to various hypothalamic, midbrain and medullary nuclei (LeDoux, 2000b). Because emotional information is processed in a serial fashion between the lateral and central nuclei in this framework, it is referred to as the serial view. However, recent evidence poses substantial problems for this serial view and its central assumptions including the need for plasticity in the

lateral amygdala or the interaction of lateral and central nuclei in performance (Maren, 2005) for fear learning have been questioned (Balleine and Killcross, 2006).

1.4.2 The 'parallel' view

In stark contrast to the serial view, an alternative model has emerged from studies in appetitive Pavlovian conditioning and reward learning in instrumental conditioning. This view posits that the BLA and CeN of the amygdala operate independently and in parallel to control distinct aspects of emotional processing: While 1) the BLA encodes associations between predictive stimuli and the specific sensory features of biologically significant events, 2) the CeN encodes the association of predictive stimuli with the general motivational and affective non-specific reinforcing properties of those events (Balleine and Killcross, 2006). An important paradigm demonstrating the influence of Pavlovian cues on instrumental behaviour that has largely contributed to this view is called Pavlovian Instrumental Transfer (PIT) (see Chapter 3 for a study of PIT). More specifically, this transfer refers to the invigoration of responding elicited by a Pavlovian cue. Two routes through which Pavlovian cues can invigorate instrumental behaviour have been substantiated. One of these, called *specific PIT*, involves the ability of a stimulus to produce an expectancy for a particular outcome by activating the memory of the unique sensory-specific features of this outcome, and hence selectively enhance the performance of responses associated with that specific reward. The other, termed *general PIT*, involves the ability of a stimulus to enhance responding as a consequence of increased motivational arousal that this stimulus has acquired through its association with reward more generally. The specific and general forms of PIT have been doubly dissociated at the level of the amygdala,

where lesions of the BLA have been shown to spare general PIT, but abolish specific PIT; whereas, CeN lesions spare specific PIT but abolish general PIT (Corbit and Balleine, 2005). This double dissociation provides evidence that the BLA and CeN can function independently, thereby lending support to the parallel claim.

For a comprehensive review describing these two theories, see Balleine and Killcross (2006).

Note that available evidence endorsing either theories of amygdala function has been confined to rodent studies as inferences have been necessarily restricted to the amygdala as a whole in human imaging studies to date on the one hand, and the temporal resolution of fMRI is too poor to allow such fine-grained differentiation between the serial and parallel views on the other hand.

1.5 THEORIES OF ASSOCIATIVE LEARNING

Theories of associative learning relate to the factors which govern association formation when two stimuli are presented concurrently. The first theory of associative learning was advanced by Thorndike over a century ago, and postulated that learning consists of association formations between stimuli and behaviours whenever a response is followed by reward. This proposal formed the basis of a number of subsequent theories of associative learning, the most influential of which are briefly described in this section.

Associative learning involves learning associations between two stimuli or between a behaviour and a stimulus. The former is called Pavlovian or classical conditioning and involves repeated presentations of a previously neutral stimulus (conditioned stimulus or CS) with a biologically significant stimulus (unconditioned stimulus or US) which

elicits a response (unconditioned response or UR). After conditioning, this response (now called conditioned response or CR) will become elicited by the CS alone. The latter is called instrumental or operant conditioning and involves either the reinforcement or punishment of a certain behaviour (also called response), resulting in an alteration of the probability that this behaviour will happen again. These two types of conditioning differ in that Pavlovian conditioning involves strengthening of the stimulus-outcome association while instrumental conditioning involves strengthening or weakening of a behaviour based on its consequences (e.g. reward or punishment).

The most influential theory of associative learning was proposed by Rescorla and Wagner (Rescorla and Wagner, 1972). In this framework, the strength of a Pavlovian CR depends on the associative strength of the CS. They accounted for this change in associative strength VA for stimulus A on a given trial, with the following equation:

$$\Delta VA = \alpha\beta(\lambda - VT)$$

The change in associative strength is directly related to the discrepancy between the magnitude of the actual outcome λ , and the associative strengths of all stimuli present on trial VT (also called prediction error). This change is modulated by two learning-rate parameters with values between 0 and 1. The value of α is determined by the salience of the CS, and that of β by characteristics of the reinforcer. The most important feature of this theory is the prediction error component, which permits to explain a very wide range of experimental findings, including blocking (Kamin, 1969) and inhibitory conditioning. However, one caveat of the Rescorla-Wagner theory is that it does not explain latent inhibition, where a simple exposure to a stimulus is enough to reduce its associability (Lubow and Moore, 1959).

Another important theory of associative learning which accounts for latent inhibition was proposed by Mackintosh (Mackintosh, 1975). In his model, the associability of a stimulus is determined by how accurately it predicts reinforcement. If the CS is the best available predictor of a US, its associability will be high, but if the stimulus is a poor predictor of reinforcement, then it will be low. The equation below shows how the associability of stimulus A, α_A influences the rate of conditioning with this stimulus:

$$\Delta V_A = \alpha_A \beta (\lambda - V_A)$$

Note that unlike in the Rescorla-Wagner equation, the change in associative strength of a stimulus is determined by the discrepancy $(\lambda - V_A)$ rather than $(\lambda - V_T)$ implying that in this model stimulus selection effects are not attributed to stimuli competing for a limited amount of associative strength. However, an important problem with Mackintosh's theory concerns the prediction that attention to a stimulus will increase if it is the best available predictor of reinforcement.

Another theory that predicts that attention to a stimulus will decrease if it is a good predictor of reinforcement was proposed by Pearce and Hall (Pearce and Hall, 1980). Indeed, in their model, attention to a stimulus is necessary while subjects are learning about its significance, but once learning has reached a stable asymptote no further attention to the stimulus is required. The associability of stimulus A on trial n, is determined by the absolute value of the discrepancy $\lambda - V_T$ for the previous occasion on which A was presented (where V_T is determined in the same way as for the Rescorla-Wagner theory):

$$\alpha_{An} = |\lambda - V_T|^{n-1}$$

The associability of a stimulus will be high when it has been followed by a US that is unexpected (when $\lambda - VT$ is high), but its associability will be low when it has been followed by a US that is expected (when $\lambda - VT$ is low).

These theories of associative learning provide formal explanations for findings from both Pavlovian and instrumental conditioning (see Chapter 5 for an implementation of the Rescorla-Wagner model in instrumental conditioning and Chapter 6 for an implementation of both Rescorla-Wagner and Pearce-Hall models in Pavlovian conditioning) and there is evidence in support of each of the three theories described in this section (Pearce and Bouton, 2001).

As will be shown in Chapter 3, Pavlovian and instrumental conditioning can interact in at least two ways, as evidenced by rodent lesion studies of Pavlovian-Instrumental transfer (or PIT). One type of PIT is called 'specific' because it activates the representation of the specific sensory features of a US. It will be manifested when a CS previously paired with a specific US will invigorate a behaviour previously leading to this same US. It is interesting to note that this type of interaction may relate to Wagner's theory of standard operating procedures (SOP) in memory, which addresses issues pertaining to internal representations of the CS and US and how these influence performance (Wagner, 1981). The other type of PIT is called 'general' because it activates representations based on a general affective motivational state. It is manifested when a CS paired with any positive US will invigorate a behaviour previously leading to any US, just because the CS has acquired motivational value through its pairing with reward generally. This type of PIT may relate to occasion setting effects in that a CS may alternatively control responding in a manner that is independent of its direct association with the US (thus the CS appears to "set the

occasion for" responding to a second CS without entering into a direct association with the US) (Holland, 1985).

2.1 FUNCTIONAL MAGNETIC RESONANCE IMAGING (fMRI)

Cognitive neuroscience has been revolutionized by the advent of fMRI, a non-invasive imaging technique for imaging both the structure and function of the human brain in vivo. One major advantage of fMRI over other neuroimaging techniques such as electroencephalography (EEG), magnetoencephalography (MEG) and positron emission tomography (PET) is its greater spatial resolution; which allows for a more accurate localization of cognitive functions in the brain than previously achievable. Note that although spatial resolution is improved with this technique, voxels, which correspond to the smallest unit, have a volume no smaller than 27mm^3 (3mm isotropic voxel size is the most commonly used in standard fMRI), which still precludes the identification of locus pertaining to small areas such as amygdalar subregions. Notwithstanding, fMRI is a remarkable tool to functionally map loci of changes in brain haemodynamics corresponding to very specific signals or examine interactions between functional areas driven by the intrinsic connectivity of the brain. Furthermore, the development of model-based fMRI has enabled neuroscientists to not only investigate where but also how various computational signals were implemented in the brain (O'Doherty et al., 2007). Incorporating computational models into fMRI studies is an important way of moving the field toward a more quantitative and theory-driven approach to experimental neuroscience. This powerful method was used in this thesis and has allowed us to pinpoint a number of important computational signals in different areas of the amygdala in the studies described in chapters 5 and 6.

fMRI is an extremely complex tool which relies on the interaction of the physics underlying magnetic resonance imaging and a chain of physiological processes inherent to neuronal function.

2.1.1 Physics of fMRI

When placed in an external magnetic field, protons, which are slightly magnetic due to the property of spin, tend to line up with this field. Consequently, water (H_2O), which is not normally magnetic (protons point in random directions), becomes magnetized in a magnetic field (Figure 2.1).

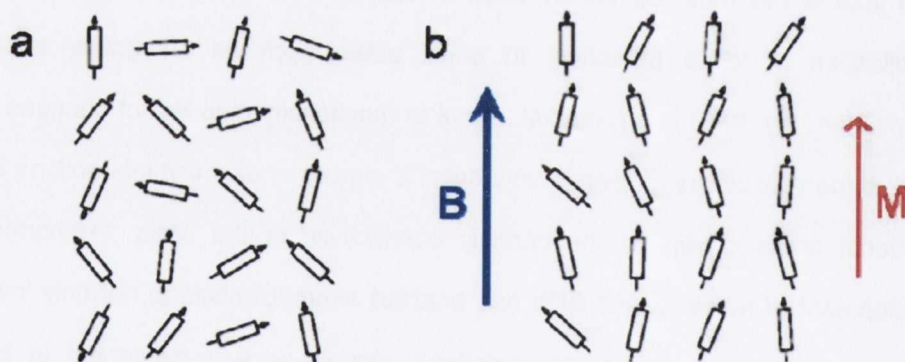


Figure 2.1. Magnetisation of water. Protons are slightly magnetic. In the absence of external magnetic force (a), the protons point in random directions, resulting in a net magnetisation of zero. However, when placed in a magnetic field **B** (b), the protons tend to align with this field, producing a net magnetisation, **M**.

When placed in the strong magnetic field of the bore of an MRI scanner (usually 1.5 or 3 Tesla), the magnetization of the water molecules of the human body lines up

with this external field. The alignment of the water proton magnetization is then disturbed by the microwave radiofrequency (RF) radiation applied. The magnetization of H₂O relaxes back to alignment with the external field when the RF is turned off, and the protons re-emit RF. A receiver in the bore detects this RF radiation, which is the basis for the MR signal. The application of magnetic gradients allows for the localization of the sources of the RF signals emitted and the frequency of the RF emitted from a specific area depends on the strength of the magnetic field at this given location. A slice-selective RF pulse is applied to excite protons in a specific slice of the participant's brain. Localization in a second direction is possible as protons along this second axis emit RF at different frequencies by applying another gradient to the measured RF signal from this slice (frequency encoding). Another gradient affecting the phase of the RF signal along a 3rd axis perpendicular to the slice selection and frequency encoding directions provides localization in the third direction (phase encoding). The location of the signal source is reconstructed by detecting the RF signal over a range of frequencies, thereby allowing the signal from every voxel (x*y*z cube from which a signal is detected) in the brain to be calculated.

The amplitude of the RF signal is correlated with the number of protons in that tissue, and thus with the density of this tissue. As a consequence, different tissue properties can be investigated by manipulating the applied RF pulse and the applied magnetic field, enabling the acquisition of images highlighting different tissue types. In this thesis, two types of images will be mentioned: T1-weighted structural images and T2*-weighted functional images.

2.1.2 Physiology of fMRI

fMRI relies on the interaction between the physics touched on above and a cascade of physiological processes. Neural activity is not detected through the electrical activity of neurons as would ideally be the case, but instead reflects a change in regional cerebral blood flow (rCBF) (Roy and Sherrington, 1890). The population of water molecules in a given area of the brain is altered by changes in local blood flow, volume and blood oxygenation, which will in turn affect the distribution of the local magnetic field and thus, the MR signal. As a consequence, neuroimagers are interested in the fluctuations in the time course of the MR signal that result from neural activity changes.

Local neural activity determines haemodynamic changes: neurons require oxygen when they fire, resulting in a difference in the concentration of oxyhaemoglobin and deoxyhaemoglobin surrounding that neuron. Oxyhaemoglobin is diamagnetic while deoxyhaemoglobin is paramagnetic and introduces inhomogeneities to the magnetic field which detract the MR signal. Therefore, it is the difference in the magnetic properties of oxygenated and deoxygenated haemoglobin that provides an indirect measure of neuronal activity.

As neurons fire, they place a high demand on oxygen, resulting in a large blood flow increase and an oversupply of oxyhaemoglobin to the area. This causes a decrease in the concentration of deoxyhaemoglobin and thus an increase in the MR signal. This effect is referred to as BOLD (blood oxygen level dependent) and is the basis for most fMRI experiments (including those of the present thesis). It has been shown that the measured BOLD response is mostly due to pre-synaptic glutamatergic activity of neurons, which is strongly correlated with post-synaptic firing rates

(Magistretti and Pellerin, 1999). An important characteristic of the BOLD response is that it is relatively sluggish compared to the associated neuronal response: the change in the MR signal occurring after neuronal stimulation takes ~1-2 seconds while it takes another 4-6 seconds for the signal to reach its maximum and between 14-18 seconds to return to baseline (see Figure 2.2).

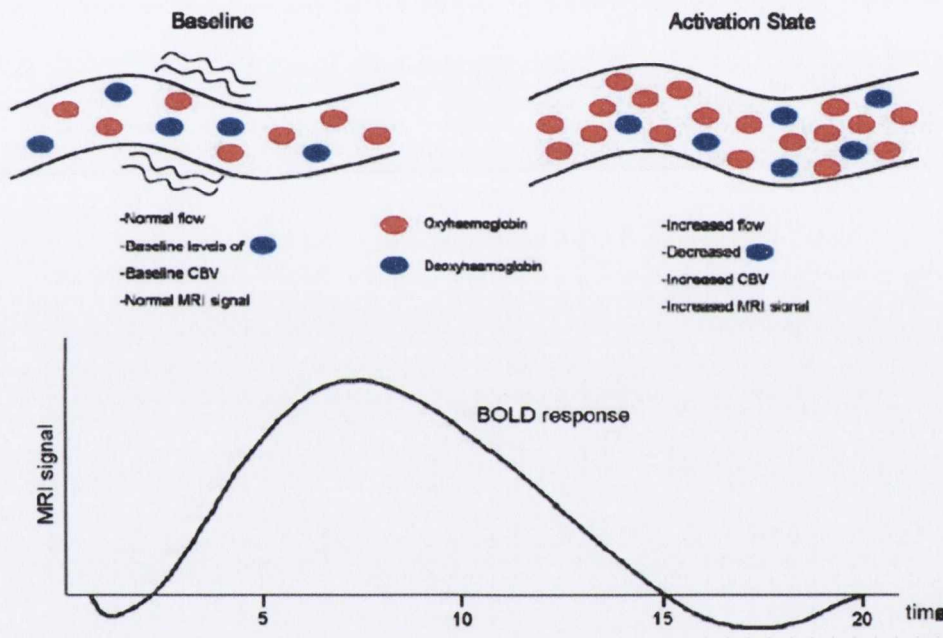


Figure 2.2. The BOLD response. In the baseline state there is distortion of the magnetic field around a blood vessel due to high concentrations of deoxyhaemoglobin. When neurons fire, there is an increase in cerebral blood volume (CBV) and blood flow resulting in an oversupply of oxyhaemoglobin. This causes an increase in the fMRI signal (top panel). The shape of the haemodynamic response is also shown (bottom panel). It takes 1-2 seconds for the MRI signal to change following neural activity, 6-10 seconds for it to reach its peak, and 14-18 second to return to baseline. The signal often displays a pre- and a post-undershoot, reflecting initial and consequent decreases in the ratio of oxy- to deoxyhaemoglobin.

At a standard resolution, it is possible to acquire enough slices to construct a whole brain image in ~2 seconds. However, at higher resolutions like the ones used in this thesis, the slices are much thinner and it is thus necessary to acquire fewer slices to remain below 2 seconds. Using analysis techniques concentrating on the BOLD signal, it is then possible to build a time course of the signal for each voxel in the brain at a 1 to 2 second temporal resolution and determine the location of neuronal activity.

2.1.3 Block, event-related and mixed designs

For each fMRI paradigm, the experimenter has to choose the optimal type of design for his/her purposes. In the early days of fMRI, block designs were mostly used due to the fact that fMRI relied on techniques developed in the field of PET. In block designs, functional images are acquired during alternating blocks of stimulus presentation ('on' periods) and some other control condition ('off' periods) which usually last ~30 seconds. The 'on' and 'off' periods are then compared, the presumption being that regions showing greater activation for the 'on' relative to the 'off' period are involved in the cognitive process elicited by the task presented during the 'on' period. This is known as the logic of 'cognitive subtraction' (Posner et al., 1988).

This block-design approach provides significant statistical power in calculating differences in activation between task conditions; however, the types of possible experiments are limited by the requirement to present stimuli of the same task condition for an entire block at a time. In addition, cognitive subtraction presumes that the transform between the neural signal and the BOLD signal is linear, which has been shown not to be completely the case (Vazquez and Noll, 1998) and which may lead to artifactual differences between conditions in the subtraction.

Event-related designs were developed in order to overcome these apparent limitations. These designs allow the detection of changes in BOLD evoked by discrete neural events associated with single behavioural trials, with the possibility for stimuli to occur randomly. For each condition, the change in BOLD observed over a number of discrete trials can be averaged in order to get a mean response in each voxel of the brain. In addition to allowing for the examination of the functional response to single trials, event-related designs allow for the separation of functional responses occurring during different temporal portions of the same trial (e.g. at the time of cue, choice and reward in decision-making tasks). Note that the potential problems arising from the nonlinearity of the BOLD response that are relevant for block designs studies also apply for event-related designs, although the error introduced by this nonlinearity is likely to be small (Buxton, 2002).

When looking for brain activity related to tonic, sustained task-related processes, and transient, phasic event-related processes within the same task, 'parsing' of functional activation can be achieved using a mixed block and event-related design (Donaldson, 2004; Visscher et al., 2003). A schematic of the different task designs is presented in figure 2.3. The tasks designed for the present thesis exclusively involve event-related designs.

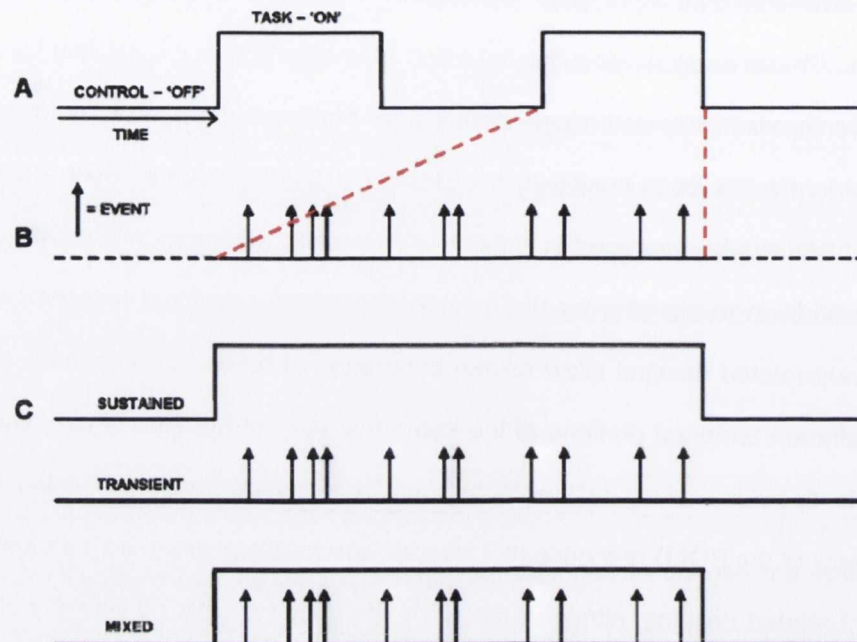


Figure 2.3. Schematic of task designs in fMRI. (A) In block designs, blocks of task performance are interspersed with blocks of control task performance or rest. (B) In event-related designs, activity related to individual trials within a task is examined. (C) In mixed designs, examination of both sustained activity corresponding to the task, and transient activity corresponding to the event-related response to specific trials within the task is enabled. Adapted from Visscher et al. (2003).

2.1.4 Issues associated with fMRI

As mentioned earlier, the main issue of fMRI is that measures are based on blood flow as opposed to directly measuring neural activity. However, an increasing body of research provides evidence for the correspondence between the findings of fMRI and electrophysiological techniques which directly measure the electrical activity of the brain (Detre et al., 1995; George et al., 1995; Puce et al., 1995) and there is also growing evidence that the blood flow response induced by neural activity correlates with

synaptic firing rates in rats, monkeys and humans (Caesar et al., 2003; Logothetis et al., 2001; Mukamel et al., 2005). The basis of fMRI research will continue being strengthened with progress in understanding the relationship between neural activity and the BOLD signal.

fMRI has received lots of criticism since its infancy, such as being compared to phrenology (Uttal, 2001). Nevertheless, purely localisationist approaches are a minority in the field, and cognition and behaviour are instead regarded as the consequence of the concerted activity of networks of functional brain regions. This is particularly evident in studies of functional and effective connectivity, investigating changes in neural interactions associated with specific signals (Buchel and Friston, 2000; Cabeza et al., 1997).

An important issue relates to the temporal resolution of fMRI, which limits the accurate separation of events occurring in close succession, meaning that even event-related fMRI has a much poorer temporal resolution than EEG or MEG. A fruitful avenue resides in the combination of the temporal resolution of these electrophysiological techniques with the greater spatial resolution of fMRI, which offers the possibility of examining the precise spatiotemporal locus of neuronal activity associated with different signals (Di Russo et al., 2003; Logothetis et al., 2001; Martinez et al., 1999; Mukamel et al., 2005; Mulert et al., 2004).

To conclude, fMRI is a widely-used, non-invasive and time-effective neuroimaging technique that offers relatively good spatial resolution compared to other available techniques to study the human brain. However, as will be thoroughly described in the next section, the spatial resolution of current fMRI studies still precludes the segregation of function within small or heterogeneous structures such as

the amygdala. Furthermore, as mentioned in the introduction of this section, the adoption of event-related fMRI in combination with computational modeling designs reveal not only 'where' but also 'how' a specific signal is implemented, rendering fMRI analysis much more powerful. Indeed, a model-based approach enables the use of fMRI data to discriminate between existing theories of brain function and to advance new theories in a way not previously achievable by conventional fMRI or behavioural studies (O'Doherty et al., 2007). The goal of this thesis was to gain a better understanding of the functional and computational role of amygdalar subregions in reward learning using increased spatial resolution by developing both acquisition techniques and analysis methods optimized for the amygdala. Therefore, throughout this thesis, an emphasis is maintained on the importance and future potential of coupling high-resolution imaging (which is described in detail in the section below) with computational modeling (described in detail in O'Doherty et al., 2007) whenever applicable.

Finally, while there are some limitations researchers should be aware of when using fMRI, the field is constantly evolving, providing an ever clearer window on the human brain.

2.2 HIGH-RESOLUTION IMAGING OF THE HUMAN AMYGDALA

As touched on above, although the spatial resolution of standard fMRI is greater than that of other available tools to study the human brain, it is nevertheless too coarse to permit making strong claims about localization of functions for different substructures of any small heterogeneous brain structure. The animal literature has undoubtedly influenced hypotheses driving studies of the human amygdala by detailing different

roles for the BLA versus CeN in associative learning (Balleine and Killcross, 2006; LeDoux, 2007); however, the limited spatial resolution of conventional neuroimaging studies has largely prevented progress toward the understanding of the role of these amygdalar subregions in humans in reward learning (and indeed in any other field). To overcome the issue of limited spatial resolution, high-resolution functional MRI (hr-fMRI) has recently been developed to afford unique leverage on the functional properties of specific brain substructures.

2.2.1 The development of high-resolution imaging

Over the past decade, high spatial resolution functional MRI has started to be implemented in order to explore subregion functions of human MTL (typically the hippocampal formation) and visual cortex in particular. It was first used by two research teams over a decade ago to examine resting states activity in the MTL using a voxel size of $0.9 \times 0.9 \times 5\text{mm}$ and a magnetic field of 1.5T (Small et al., 2000a; Small et al., 2000b) and novelty encoding effects in the MTL using a voxel size of $1.6 \times 1.6 \times 3\text{mm}$ and a magnetic field of 3T (Zeineh et al., 2000). Since these pioneering reports of hr-fMRI, relatively high-resolution functional and structural volumes have been acquired, which has permitted the segmentation of anatomical Region of Interests (ROIs) on the latter in neuroimaging studies investigating the MTL. In hr-fMRI of the MTL, this segmentation is usually guided by anatomical landmarks correlating with MTL cytoarchitectonics available in MTL atlases (Duvernoy, 2005; Insausti, 2004) and standards developed in structural MRI studies of the MTL (Insausti et al., 1998; Pruessner et al., 2002; Pruessner et al., 2000). Component hippocampal subfields and surrounding cortical structures are now routinely segmented into the entorhinal cortex,

perirhinal cortex, subiculum, CA1, and a combined region consisting of the dentate gyrus, CA2, and CA3 (areas which are still not unambiguously distinguishable using current techniques) (Carr et al., 2010) (Figure 2.4).

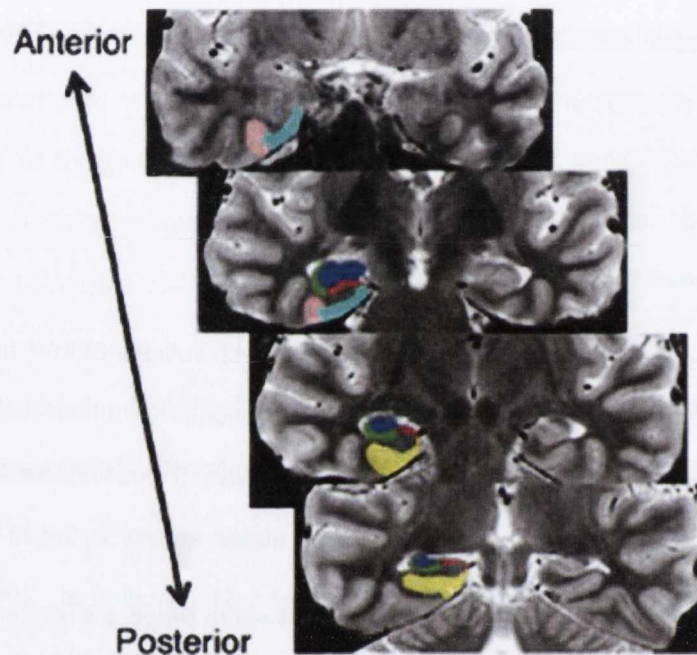


Figure 2.4. High-resolution anatomical images of human MTL displaying manually drawn hippocampal ROIs: CA1 (green), DG/CA2/3 (blue), and the subiculum (red), as well as PRC (pink), ERC (cyan), and PHC (yellow). Adapted from Carr et al. (2010).

In hr-fMRI studies investigating small areas of the visual cortex such as ocular dominance columns or V1, much higher resolutions of $0.47 \times 0.47 \times 3\text{mm}$ and $0.75 \times 0.75 \times 3\text{mm}$ respectively have been used at a strong magnetic field of 4T (Cheng et al., 2001; Sun et al., 2007). More recently, hr-fMRI protocols have been used to study memory aspects of the MTL using a 1.5mm isotropic voxel size and a magnetic field of

3T (Bakker et al., 2008; Kirwan et al., 2007), the properties of the fusiform face area using a 1mm isotropic voxel size at a magnetic field of 3T (Grill-Spector et al., 2006), dopaminergic signals in the human VTA at a magnetic field of 3T (D'Ardenne et al., 2008) and midbrain anatomy at a high field strength of 7T (Eapen et al., 2011). Hr-fMRI is developing fast and spreading to various areas of the brain and has already proven to be highly valuable, at least at the level of the hippocampal formation (see Carr et al. (2010) for a comprehensive review of hr-fMRI in the MTL).

2.2.2 The amygdala: a challenging structure to image

Progress in understanding the role of amygdalar subregions has been hindered by the difficulties in probing this small subcortical structure buried deep within the MTL. In this section, aspects of the difficulty in scanning the amygdala will be described; however, the human amygdala is not only challenging to image, but to study in general with currently available tools, as will be discussed in the section 6.3 of the General Discussion (Chapter 7).

An important caveat concerning the amygdala is that signal quality is not uniform across the whole complex, in that the anterior and medial part of the ventral amygdala yields a lower signal to noise ratio (SNR). These inhomogeneities produce field distortions that can lead to a substantial and frequently irrecoverable loss of signal in this region during echo planar imaging, as well as signal mislocalization (Cordes et al., 2000; Farzaneh et al., 1990; LaBar et al., 2001). This magnetic-susceptibility-induced signal loss is due to the fact that medially, the amygdala is bounded by sinuses and ventrocaudolaterally by the lateral ventricle. This contributes to inter-participant and inter-hemispheric variability in amygdala activation. Furthermore, these susceptibility

artifacts may lead to false negative as well as false positive results (although the bias is toward false negative findings due to signal dropout).

Another issue relates to variability in the exact location of the amygdala (and amygdalar subregions as well) across participants. Although the general location of the amygdala appears fairly consistent as reflected by the similarity in its coordinates in different brain atlases (Mai et al., 2008; Talairach, 1988), the amygdala possesses an average width of around 15mm and a slightly longer length which makes it a small structure, hence prone to misregistration across participants. Furthermore, standard normalization procedures are extremely accurate for cerebral structures (for which they have been designed), but perform rather poorly for small subcortical structures like the amygdala. Obviously, this issue is critical for voxel-wise statistical analyses which require precise spatial registration across participants.

Finally, localization error can be introduced by draining vein effects which may mislocalize foci by 5mm or more (i.e. one third of the amygdala width) and adds uncertainty to findings arising in fMRI studies (Kinahan and Noll, 1999; Lai et al., 1993).

Therefore, even though the amygdala can relatively easily be identified anatomically, obtaining reliable hemodynamic signal from this region can be difficult. Nevertheless, these problems of SNR, inhomogeneities, anatomic variability and mislocalization can be partially alleviated or overcome.

2.2.3 A recipe for high-resolution imaging of the amygdala

An important stage of this PhD has been to develop an hr-fMRI protocol in order to address questions pertaining to the functional contributions of BLA and CeN to

specific aspects of reward and associative learning. Given the difficulties in imaging the amygdala as a whole, getting reliable hemodynamic signals from different amygdalar subregions presents many challenges. Increasing the spatial resolution of fMRI is not just about decreasing the voxel size of the functional data but requires a variety of resolution preserving data analysis procedures. To fulfill our goal, particular attention was paid to the voxel size, the unwarping of functional data, the need of specifically normalizing the amygdala, the importance of not applying any extrinsic spatial smoothing and of modeling out the physiological fluctuations (see figure 2.5 for an illustration of our hr-fMRI protocol). Using this hr-fMRI protocol, not only has it been possible to address all of the issues mentioned above to improve signal quality from the amygdala as a whole, but we have been able to considerably increase the spatial resolution of our data, thereby allowing us to extract reliable BOLD signal from different amygdalar subregions.

2.2.3.1 Decrease your voxel size

High-resolution fMRI usually refers to the acquisition of functional images at a smaller voxel size (i.e. higher spatial resolution) than is usually acquired. Typically, the in-plane resolution of hr-fMRI voxels is smaller than $2 \times 2 \text{ mm}^2$ while that of standard fMRI voxels is usually larger than $3 \times 3 \text{ mm}^2$. Ideally, the slice thickness is the same size as the in-plane resolution, thus creating an isotropic voxel size (i.e. the x, y and z dimensions are identical); however, the slice thickness being crucial for SNR as it decreases proportionally to the slice thickness, in some instances, it will have to be slightly larger than the in-plane resolution to avoid losing too much signal. In the current thesis, different voxel sizes have been used: 1) an anisotropic voxel size of 1.58×1.63

x 2.2 mm^3 on the Philips 3T scanner equipped with an 8-channel coil at Trinity College Dublin, and 2) an isotropic voxel size of $1.8 \times 1.8 \times 1.8 \text{ mm}^3$ on the Siemens 3T scanner equipped with a 32-channel coil at the California Institute of Technology. These sizes result in a volume of less than 6 mm^3 per voxel in both cases which is more than 4 times smaller than the usual voxel volume of $3 \times 3 \times 3 = 27 \text{ mm}^3$. Higher spatial resolution also implies acquiring many more voxels and slices, which leads to huge datasets to analyze. Therefore, to prevent this and to keep volume acquisition at a TR of $\sim 2\text{s}$, researchers generally only acquire a number of slices centered on a specific portion of the brain which includes their area of interest (resulting in a smaller field of view than typically acquired).

2.2.3.2 Unwarp your data

Although only indirectly linked to resolution per se, the problem of spatial EPI distortions needs to be addressed to prevent signal mislocalization, especially in the amygdala which encompasses the anterior MTL, an area containing high levels of inhomogeneity in magnetic susceptibility due to its proximity to the sphenoid sinus. The field distortions produced by these inhomogeneities can be substantially attenuated by unwarping (i.e. correcting for the measured displacement) the functional data using gradient field maps (or B_0 maps) acquired during scanning when preprocessing the functional data, which allows for increased precision in functional localization.

2.2.3.3 *Specifically normalize the amygdalae*

Another methodological issue limiting the SNR in group analyses that requires particular attention is inter-individual anatomic variability (Insausti et al., 1998). Atlas-based approaches used to register whole-brain EPI images across participants look for a global optimum alignment of the cortex which is achieved under the limitations imposed by the available degrees of freedom, and which is at the expense of regional accuracy of small subcortical structures. Consequently, the amygdala becomes particularly vulnerable to misregistration across participants and BOLD signal in this region may be underestimated or possibly missed (Miller et al., 2005). Therefore, a specific normalization of the amygdala becomes crucial when performing voxel-level group analyses of hr-fMRI data. Alignment (or registration) of the amygdala is substantially improved by a Region of interest alignment (ROI-AL) approach, in which segmentations of ROIs are drawn on structural images and aligned directly, resulting in an increased statistical power (Yassa and Stark, 2009). This normalization involves three steps. First, the amygdala needs to be manually segmented (preferably by a single observer to avoid an observer bias) on each coronal image containing the amygdala for each participant's structural scan. This segmentation needs to be done using detailed tracing guidelines based on an atlas of the human brain (Mai et al., 2008). Guidelines could read as follows: In its rostral part, the amygdala is bordered ventromedially by the entorhinal cortex, ventrally by the temporal horn of the lateral ventricle and subamygdaloid white matter and laterally by white matter of the temporal lobe. Midrostrocaudally, the amygdala increases in size and is bordered ventromedially by a thin tract of white matter separating the amygdala and the entorhinal cortex, laterally by the white matter of the temporal lobe and medially by the semiannular sulcus. Caudally, the amygdala is bordered dorsally by the substantia innominata and

fibers of the anterior commissure, laterally by the putamen, ventrally by the temporal horn of the lateral ventricle and the alveus of the hippocampus and medially by the optic tract. These segmented amygdalae will be used to create a template amygdala. The second step of this normalization procedure consists in wrapping or aligning each participant's amygdala to match this template amygdala. Recent advances in cross participant registration techniques such as fully deformable nonlinear registration algorithms (Ekstrom et al., 2009; Yassa and Stark, 2009) have proven to be exceptionally accurate in the alignment of hippocampal subfields (see: <http://darwin.bio.uci.edu/~cestark/roial/roial.html>). The registration to the template amygdalae can be done using the last iteration of these alignment tools which is called ROI-Demons. This technique uses Thirion's original demons algorithm as implemented by Vercauteren and which enforces smooth deformations by operating on a diffeomorphic space of displacement fields (Thirion, 1998; Vercauteren et al., 2007). When wrapping the participant's amygdalae to the template, displacement fields (or transformation calculations) are created. The last step of this normalization consists in applying these individual displacement fields to each subject's normalized EPI scans in order to specifically normalize their amygdalae to the template amygdalae. This whole procedure affords reliable voxel-level evaluation of hr-fMRI amygdala activity at the group level, allowing activation patterns consistent across participants to be readily visualized.

2.2.3.4 Do not smooth your data

Importantly, high-resolution can be preserved by forgoing the application of a spatial smoothing kernel. Most hr-fMRI studies have avoided applying extrinsic

smoothing while others (Boll et al., 2011; Gamer et al., 2010) have applied minimal smoothing in an effort to boost the SNR for voxel-level analyses. Obviously, this is a critical point that will determine the fidelity of signal localization.

2.2.3.5 Remove physiological fluctuations from your data

Finally, when doing hr-fMRI, physiological fluctuations become problematic and need to be accounted for. Participants' cardiac and respiratory signals can be easily recorded with a pulse oximeter placed on the forefinger and a pressure sensor placed on the umbilical region during scanning. These signals can then be removed from time-series images during data analysis using the RETROICOR algorithm which estimates thirteen regressors of no interest (4 related to heart rate, 9 related to respiration) that are included in the general linear model (GLM) of each participant (Glover et al., 2000).

2.2.3.6 Outline your ROIs

In order to contrast the functional roles of BLA and CeN in the studies presented in this thesis, we used a manual segmentation approach to delineate the amygdala substructures as follows:

Amygdalar subdivisions were hand-drawn on our template amygdalae using the Atlas of the Human Brain (Mai et al., 2008). We delineated three sub-areas within the amygdala: the BLA complex comprised of the basomedial, basolateral and lateral nuclei; the CeN complex comprised of the central and medial nuclei; and the cortical complex (or cortical nucleus). In its most rostral part, the amygdala is exclusively composed of the BLA. The cortical nucleus appears in the dorso-medial part of mid-

rostral amygdala. The CeN appears slightly more caudally than the cortical nucleus in the most dorsal part of the amygdala. The BLA increases in size as one moves caudally from the anterior amygdala, has its maximal size midrostrocaudally and then decreases as one moves further back toward the caudal amygdala, whereas the cortical nucleus and CeN slightly enlarge midrostrocaudally, but do not decrease in size as one moves further caudally within the amygdala. The cortical nucleus ends midcaudally, the BLA ends in caudal amygdala while the CeN ends in the most caudal part of amygdala.

In conclusion, enhanced spatial resolution can be achieved with resolution preserving data analysis procedures as the ones described above and allows for localization of BOLD in specific areas of the amygdala, with activation in a given voxel being an indirect hemodynamic proxy for activity from several thousands of underlying neurons (Logothetis, 2008). Note that we are the first to use all of these techniques within the same study, although prior hr-fMRI studies have used one or two of these techniques. For instance, Gamer et al. (2010) published an hr-fMRI study investigating the role of amygdalar subregions in gaze and attention; however, they only used a relatively small voxel size and applied a 4mm kernel smoothing which is smaller than non-hr-fMRI studies, but much larger than the studies described in this thesis (which completely forwent the application of a spatial smoothing kernel).

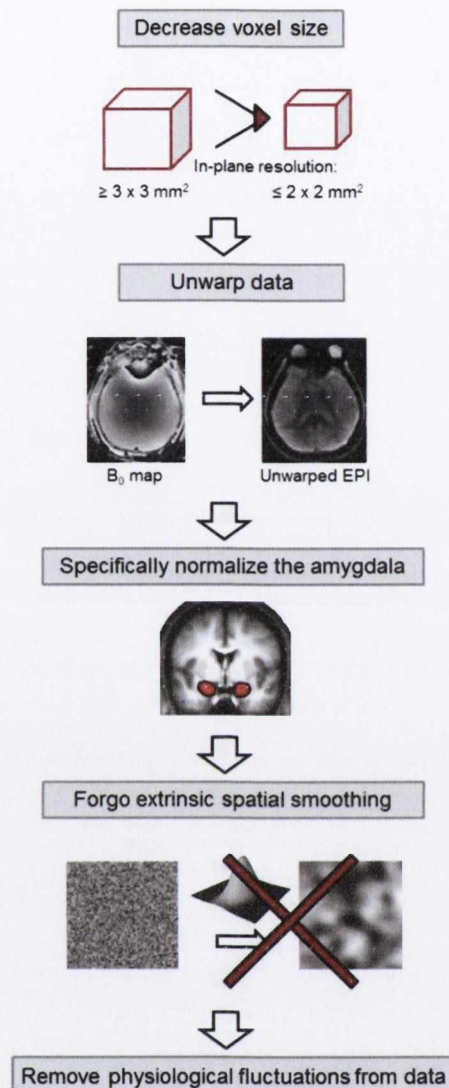


Figure 2.5. Illustration of our hr-fMRI protocol. We developed an fMRI sequence using a small voxel size (volume more than four times smaller than in standard fMRI). The data was unwarped to remove spatial EPI distortions which are prominent in the amygdala due its close proximity to sinuses. Participant's amygdalae underwent a specific normalization procedure to reduce inter-participant anatomic variability. No extrinsic spatial smoothing was applied to the data to ensure the best fidelity possible in signal localization. Finally, participants' cardiac and respiratory fluctuations were removed from time-series during data analysis.

CHAPTER 3 NEURAL CORRELATES OF SPECIFIC AND GENERAL PAVLOVIAN TO INSTRUMENTAL TRANSFER WITHIN HUMAN AMYGDALAR SUBREGIONS: A HIGH-RESOLUTION fMRI STUDY

It is widely held that the interaction between instrumental and Pavlovian conditioning induces powerful motivational biases. Pavlovian-Instrumental Transfer (PIT) is one of the key paradigms demonstrating this effect, which can further be decomposed into a general and specific component. Although these two forms of PIT have been studied at the level of amygdalar subregions in rodents, it is still unknown whether they involve different areas of the human amygdala. Using a high-resolution fMRI (hr-fMRI) protocol optimized for the amygdala in combination with a novel free operant task designed to elicit effects of both general and specific PIT, we demonstrate that a region of ventral amygdala within the boundaries of the basolateral complex and the ventrolateral putamen are involved in specific PIT, while a region of dorsal amygdala within the boundaries of the centromedial complex is involved in general PIT. These results add to a burgeoning literature indicating different functional contributions for these different amygdalar subregions in reward-processing and motivation.

3.1 INTRODUCTION

Pavlovian and instrumental processes interact behaviourally in at least two different ways. In specific Pavlovian to instrumental transfer (specific PIT), a Pavlovian stimulus associated with a particular outcome selectively enhances the performance of

responses associated with that specific outcome. In general PIT, a Pavlovian cue engenders a general increase in the vigor of instrumental responding, independently of the identity of the specific outcome involved (Balleine and Killcross, 2006; Dickinson and Balleine, 2002; Holland, 2004). For instance, imagine that while shopping on your way to a French restaurant, Snickers bars, are advertised in the store. An example of manifestation of specific PIT would be for you to go and buy a Snickers bar while general PIT would be manifested if the sight of the chocolate bar provided a stronger drive for you to run to the restaurant. These two different forms of PIT have been suggested to reflect the manifestation of two very distinct forms of incentive processing: a sensory specific system involved in encoding emotional events based on their specific sensory properties and a general system in which the affective significance of an outcome is encoded irrespective of the specific sensory properties of that outcome (Balleine and Killcross, 2006). In spite of the importance of these mechanisms in the motivational processes influencing human choice, very little is known about the neural mechanisms underpinning them, particularly in the human brain.

Animal studies investigating the neural underpinnings of PIT have focused to a considerable extent on the role of distinct subregions within the amygdala and ventral striatum, particularly the basolateral complex (BLA) and the centromedial complex (CeN) for the former and the nucleus accumbens core and shell for the latter. Some studies have implicated the BLA (Blundell et al., 2001), while others found that lesions of the CeN, but not the BLA abolished this phenomenon (Blundell et al., 2001; Holland and Gallagher, 2003). Corbit and Balleine (2005) attempted to reconcile these earlier findings by suggesting that the differences reported in those studies were due to differences in the form of PIT investigated. By examining the specific and general forms of PIT within the same paradigm they found that lesions of the BLA spared general PIT

but abolished specific PIT whereas CeN lesions spared specific PIT but abolished general PIT. Similarly, it was shown that lesions of the NAC core but not shell abolished PIT (Hall et al., 2001) and in stark contrast, that lesions of the NAC shell but not core abolished PIT (Corbit et al., 2001). Again, these findings were reconciled by investigating specific versus general PIT in a recent study revealing that NAC shell but not core lesions abolished the former whereas lesions of NAC core but not shell eliminated general PIT (Corbit and Balleine, 2011).

In contrast, very little is known about the neural mechanisms underpinning PIT in humans. Talmi et al. (2008) reported activity in the striatum and the amygdala in response to global PIT (the task was not designed to allow discrimination between specific or general PIT, hence the term "global" PIT), while Bray et al. (2008) reported a contribution of the ventrolateral putamen but not the amygdala in specific PIT (Bray et al., 2008; Talmi et al., 2008). However, in both of these studies, the use of a conventional imaging protocol precluded segregation of function within the amygdala, and Talmi et al.'s study was not designed to simultaneously measure specific and general PIT while Bray et al.'s study reported specific PIT but failed to report general PIT.

The goal of the present study was to investigate the role of human amygdalar subregions and ventral striatum in specific and general PIT with fMRI. We used a novel free operant paradigm adapted to the fMRI scanner to enable us to measure general as well as specific PIT within the same participants in a manner directly analogous to that done in rodent studies. A further critical innovation in the present study is the use of a high-resolution fMRI protocol (hr-fMRI) in order to enable segregation of activity within different subregions in the amygdala. We hypothesized that the ventral striatum would contribute to specific PIT as shown in previous fMRI studies, and that the BLA would

contribute to specific PIT, while the CeN of the amygdala would contribute to general PIT.

3.2 MATERIAL AND METHODS

3.2.1 Participants

Twenty-six healthy (5 females), right-handed participants with a mean age of 27.04 ± 6.09 took part in this study. All participants were free of neurological or psychiatric disorders and had normal or correct-to-normal vision. As the experiment involved food consumption, the eating attitudes test (EAT-26; see (Garner et al., 1982) was administered prior to the experiment, which indicated no eating disorders in any of the participants (score: 6.19 ± 5.49). Participants were asked to fast for at least 4 hours prior to the experiment to ensure motivation for food reward, although they were allowed to drink water. Written informed consent was obtained from all participants, and the study was approved by the Caltech Institutional Review Board.

3.2.2. Stimuli

Visual stimuli were presented via a projector positioned at the back of the room. Participants viewed a reflection of the projected image in a mirror attached to the scanner head coil. The food rewards consisted of sweet and salty snack foods (Chips Ahoy® cookies, Doritos® ranch chips, Famous Amos® cookies, Ghiradelli® milk chocolate bar, Hostess HoHos®, KitKat® candy bar, Lindt® swiss chocolate bar, MrsFields® chocolate chip cookies, Oreo® cookies, RiceKrispy® Treat, Milano® cookies, Keebler® Fudge Stripes® cookies, Keebler® Rainbow® cookies, Toblerone® chocolate bar, Twix® caramel and chocolate bar). Participants were asked to provide subjective pleasantness ratings for each of the 15 different snack foods at the time of

recruiting, and the experimenter selected three highly and equally valued food items for each participant based on their subjective ratings. Stimulus presentation and behavioural data acquisition were implemented in Matlab (The Mathworks Inc., Natick, MA, USA) with the Cogent 2000 toolbox (Wellcome Department of Imaging Neuroscience, London, UK).

3.2.3 Training

Before being scanned, to familiarize participants with the set-up, they underwent a brief training on the instrumental and Pavlovian phases.

3.2.3.1 Instrumental training

During instrumental training trials, participants were instructed that two gray squares would light up signaling which two of three possible keys were available on that trial (**Fig. 3.1a**). They were asked to press any given key multiple times in order to get rewarded with food. Participants were explicitly told to press only one of the two available responses at a time. Every time one of these two keys was pressed, it flashed in white for 50 ms. In this free-operant task, each response was rewarded on a random ratio (RR) schedule of 0.1 as in Corbit and Balleine (2005), so that a food picture was displayed once every time a key was pressed a random number of times between 5 and 15 times (i.e. hence once every 10 times on average). Each of the three responses earned one of the three distinct food rewards (**Table 3.1**, top panel). As described in Table 3.1, there were three types of instrumental trials. Given that participants were told that rewards were contingent on their presses, key presses were also recorded during outcome presentation. Depending on their number of presses per response, participants could get more than one outcome per trial in this instrumental training. Each of the three responses earned the three distinct food rewards (**Table 3.1**, top

panel). Whenever a certain food was depicted on the screen, the experimenter placed a small piece of food corresponding to the picture in the participant's left hand to be consumed immediately. Each trial lasted for 6 s and was followed by a jittered 2-6 s inter-trial interval (ITI). A trial duration of 6 seconds was achieved to allow enough time for participants to generate a large number of presses, without making it effortful. This duration was fixed to allow reliable comparisons in the number of presses across the different trial types. Food pictures were presented for 1 s. The location of the food pictures was the same across all phases for a given participant but was counterbalanced across subjects so that it appeared on the left side of the screen for half of the subjects and on the right side of the screen for the other half of the subjects. This short instrumental training was comprised of 6 trials (2 of each type). The trial order was randomized within this phase as well as across participants.

a Instrumental trials



Figure 3.1a. Instrumental training. During the instrumental trials, two of the three squares at the bottom of the screen changed color from black to gray for 6 s during which time participants were instructed to press any given button multiple times in order to get rewarded with food. The food picture corresponding to their response was displayed on one side of the screen for 1 s every 10 presses on average.

3.2.3.2 Pavlovian training

Pavlovian training trials consisted of the presentation of associations between fractal images (conditioned stimuli, CSs) and outcomes, three of which were rewarding (food rewards) and one of which was neutral (no reward) (**Fig. 3.1b**, **Table 3.1**, middle panel). There were three types of Pavlovian trials: stimulus 1 (S1) leading to outcome 1 (O1), S2 leading to O2, S3 leading to O3 and S4 leading to no outcome. Therefore, participants experienced four different outcomes during Pavlovian training, while they experienced only three in the instrumental trials. CSs were presented on one side of the screen for 6 s and rewards were delivered concomitantly during the last second of the trial (between 5 and 6 s) in a deterministic fashion. Again, every time a food picture was depicted on the screen, the experimenter placed a piece of food corresponding to the picture in the participant's left hand to be consumed immediately. The trial ended with an ITI of 2-6 s. The location of the fractal cues and food pictures was counterbalanced across participants. The assignment of the associations between fractal images and outcomes was randomized across participants. This short Pavlovian training was comprised of 4 trials (one of each type). The trial order was randomized within this phase as well as across participants.



Figure 3.1b. Pavlovian training. In the Pavlovian trials, a conditioned stimulus was presented on one side of the screen for 6 s, and the food picture corresponding to the

stimulus was displayed on the other side of the screen during the last second of the trial.

3.2.4 Task description

The scanned portion of the experiment consisted of an instrumental, a Pavlovian and a Pavlovian-Instrumental Transfer (PIT) phase (**Fig. 3.1**).

3.2.4.1 Instrumental phase

This phase was exactly the same as during the training except that participants were told that whenever they saw a food picture, the experimenter would place the corresponding food in a cup which they could retrieve at the end of the experiment. The three black squares at the bottom of the screen corresponded to the three buttons on a response box (Current Designs) that the participants held in their right hand. Each trial type was presented 10 times, for a total of 30 trials and a total duration of 5 min (**Table 3.1**, top panel).

3.2.4.2 Pavlovian phase

Again, this phase was exactly the same as during the training except that participants were told that whenever they saw a food picture, the experimenter would place the corresponding food in a cup which they could retrieve at the end of the experiment. Each trial type was presented 10 times, for a total of 40 trials and a total duration of ~7 min (**Table 3.1**, middle panel).

3.2.4.3 PIT phase

After the instrumental and Pavlovian phases, participants performed a series of transfer trials. During these trials, one of the Pavlovian cues was presented simultaneously with

the two instrumental response options (**Fig. 3.1c**) and, as in instrumental trials, participants were given the opportunity to freely respond to either of the two available options. They were explicitly told that they would not be able to see any food picture during this phase, because the area where they were usually displayed would be hidden, but that they should assume that all reward deliveries would be as they had been during the previous phases. The reason for performing this phase in extinction (i.e. no outcomes delivered) was to allow assessment of the influence of the Pavlovian cues on instrumental responding without the confounding effects of the outcomes themselves (Bray et al., 2008). As previously described, PIT effects can be further decomposed into outcome-specific and outcome-general PIT (Corbit and Balleine, 2005; Dickinson and Balleine, 2002; Holland, 2004). In specific PIT, the Pavlovian cue associated with a particular outcome enhances only an instrumental response associated with the same appetitive outcome. However, in general PIT, the Pavlovian cue enhances any instrumental response, even if the instrumental response is paired with a different appetitive outcome. Because this study was aimed at comparing the neural structures involved in these two distinct processes, we designed a task allowing us to behaviourally distinguish these phenomena. Thus, there were three different types of trials: specific, general and neutral trials (**Table 3.1**, bottom panel). To test for specific PIT, participants were given the opportunity to respond to two available responses previously paired with two particular reward outcomes, one of which also previously paired with the concurrently presented Pavlovian cue. For instance, response 1 (R1) previously paired with outcome 1 (O1) and response 2 (R2) previously paired with outcome 2 (O2) were made available in presence of stimulus 1 (S1) previously paired with O1 or stimulus 2 (S2), paired with O2. There were six subtypes of specific trials: R1 and R2 presented with S1, R1 and R3 presented with S1, R1 and R2 presented with S2, R2 and R3 presented with S2, R1 and R3 presented with S3 and R2 and R3

presented with S3. Evidence for specific PIT would be seen if the presence of the Pavlovian cue induced a higher response rate for the action associated with the same outcome as that Pavlovian cue. In the subsequent analysis, we pooled over all six specific trial subtypes, but differentiated between trials in which participants responded more to actions compatible with the Pavlovian outcome and those trials in which participants responded more to the action that was not compatible (as in Bray et al., 2008). There were 5 trials of each subtype, for a total of 30 specific trials. To test for general PIT, we also included general and neutral trials. In general trials, participants were given the opportunity to respond to two available responses, but this time the Pavlovian cue was previously associated with a different outcome that was not compatible with either response option. There were three subtypes of general trials: R2 and R3 presented with S1, R1 and R3 presented with S2 and R1 and R2 presented with S3. There were 10 trials of each subtype, for a total of 30 general trials. In neutral trials, participants were again given the opportunity to respond to two available responses but the Pavlovian cue was that associated with the neutral outcome. There were three subtypes of neutral trials: R1 and R2 presented with S4, R1 and R3 presented with S4 and R2 and R3 presented with S4. There were 10 trials of each subtype, for a total of 30 neutral trials. Evidence for general PIT would be seen if participants responded more to general than neutral trials, indicating that reward cues exert some general motivation over any instrumental response, acquired through their pairing with a reward outcome (as opposed to the neutral cue which has not been paired with a reward outcome and as such, has not acquired any motivational value). Each type of trial was presented 30 times for a total of 90 trials and a total duration of 15 min. The trial order was randomized within this phase as well as across participants.

C PIT trials

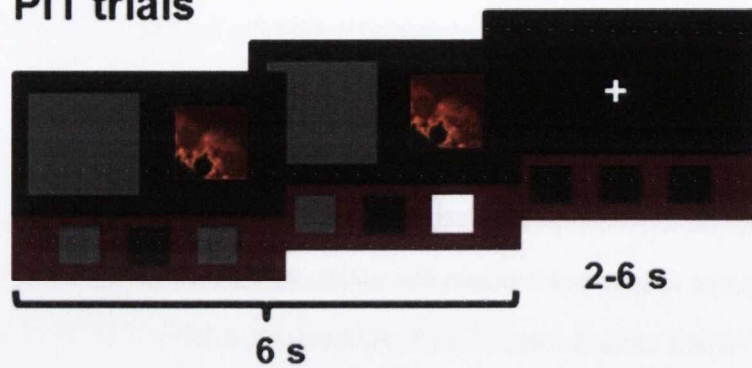


Figure 3.1c. PIT task illustration. During the PIT trials, the conditioned stimulus was presented simultaneously with two squares changing color for 6 s. Participants were instructed to press any given key multiple times in order to get rewarded with food, as in the instrumental trials. In this phase, the area where the food picture was usually displayed was hidden behind a gray square.

Phase	Trial	Pavlovian	Instrumental		Outcome	
Ins	10	-	R1	R2	O1	O2
		-	R1	R3	O1	O3
		-	R2	R3	O2	O3
Pav	10	S1	-	-	O1	
		S2	-	-	O2	
		S3	-	-	O3	
		S4	-	-		
PIT	5	S1	R1	R2		
		S1	R1	R3		
		S2	R1	R2		
		S2	R2	R3		
		S3	R1	R3		
		S3	R2	R3		
	10	S1	R2	R3		
		S2	R1	R3		
		S3	R1	R2		
	10	S4	R1	R2		
		S4	R1	R3		
		S4	R2	R3		

Table 3.1. Trial composition for the instrumental, Pavlovian and PIT phases. Ins, instrumental; Pav, Pavlovian, S1-S4, conditioned stimuli; R1-R3, three button response actions; O1-O3, food rewards.

3.2.5 Behavioural measures

3.2.5.1 Physiological fluctuations

Although both heart rates and breathing rates were recorded during scanning to be further removed from time-series (see fMRI data analysis), only the time course derived from estimates of heart rate was used as a physiological index of conditioning.

3.2.5.2 *Response rates*

Response rates were recorded both during the instrumental and PIT phases (**Fig 3.2**).

3.2.5.3 *Affective evaluations of food pictures*

Before the start of the training and before and after the PIT phase, participants were asked to rate the pleasantness of the food pictures by moving a cursor along a scale from -5 to +5.

3.2.6 **Data Acquisition**

All magnetic resonance imaging was performed on a 3 Tesla Magnetom Tim Trio scanner (Siemens Medical Solutions, Erlangen, Germany) equipped with a 32-channel receive-only phased array head coil. Since the focus of our study was on the amygdala, we acquired oblique axial T2*-weighted echo planar images (EPI) with partial brain coverage centered on the medial temporal cortex while participants were performing the task. These images also encompassed the ventral striatum, ventral part of the prefrontal cortex, the insula, the hippocampus, the ventral part of the occipital lobe and the upper part of the cerebellum (amongst other regions). Twenty-two EPI slices with ascending interleaved ordering were acquired, with an isotropic voxel size of 1.8 mm (no slice gap); thereby providing ~4.6 times the volume resolution of a typical standard 3 mm isotropic voxel size acquisition. Other imaging parameters included: repetition time (TR) = 2000 ms; echo time (TE) = 41 ms; field of view = 180 x 180 x 39.6 mm; sampling matrix = 100 x 100 voxels). We discarded the first 3 EPI volumes before data processing and statistical analysis to minimize magnetization equilibration effects on the

BOLD timeseries. Whole-brain high-resolution T1-weighted structural images (two repetitions, isotropic voxel size = 1 mm) and whole-brain T2*-weighted images (three repetitions) were acquired for each participant. To address the problem of spatial EPI distortions which are particularly prominent in the medial temporal lobe (MTL) and especially in the amygdala, we also acquired dual gradient echo B_0 field maps. Finally, to account for the effects of physiological noise in the fMRI data, participants' cardiac and respiratory signals were recorded using the MRI system's pulse oximeter and respiratory bellows.

3.2.7 Preprocessing

All EPI volumes (partial scans acquired while participants were performing the task and the three whole-brain functional scans acquired prior to the experiment) were corrected for differences in slice acquisition and spatially realigned. The mean whole-brain EPI was co-registered with the T1-weighted structural image, and subsequently, all the 'partial' volumes were co-registered with the registered mean whole-brain EPI image. 'Partial' volumes were then unwrapped using the gradient field maps. After the structural scan was normalized to a standard T1 template, the same transformation was applied to all the 'partial' volumes with a resampled voxel size of 1x1x1 mm. In order to maximize the spatial resolution of our data, no spatial smoothing kernel was applied to the data. These preprocessing steps were performed using the statistical parametric mapping software SPM5 (Wellcome Department of Imaging Neuroscience, London, UK).

3.2.7.1 Amygdalae segmentation

Amygdalae Regions of Interest (ROIs) were manually segmented for each participant by a single observer using a pen tablet (Wacom Intuos3 Graphics Tablet) in FSL View (FSL 4.1.2). This program allows magnification and the simultaneous viewing of volumes in coronal, sagittal and horizontal orientations. Amygdalae were manually outlined on each coronal image containing these regions using detailed tracing guidelines based on the Atlas of the Human Brain (Mai et al., 2008). Outlines were checked in horizontal and sagittal planes when they proved more valuable for the identification of structure boundaries. The anterior limit of the amygdala was defined using the horizontal and sagittal planes. The following guidelines were used: In its rostral part, the amygdala is bordered ventromedially by the entorhinal cortex, ventrally by the temporal horn of the lateral ventricle and subamygdaloid white matter and laterally by white matter of the temporal lobe. Midrostrocaudally, the amygdala increases in size and is bordered ventromedially by a thin tract of white matter separating the amygdala and the entorhinal cortex, laterally by the white matter of the temporal lobe and medially by the semiannular sulcus. Caudally, the amygdala is bordered dorsally by the substantia innominata and fibers of the anterior commissure, laterally by the putamen, ventrally by the temporal horn of the lateral ventricle and the alveus of the hippocampus and medially by the optic tract.

3.2.7.2 Amygdalae normalization

We normalized our functional data as opposed to extracting signal from individually segmented ROIs because performing voxel-wise inferences at the group level offers a considerably greater statistical power than individual subject ROI analyses.

Because structures in the MTL exhibit significant inter-individual anatomic variability, the signal-to-noise ratio in group analyses is substantially limited in these areas (Insausti et al., 1998). Atlas-based approaches used to register whole-brain EPI images across participants (such as SPM) look for a global optimum alignment which is achieved under the limitations imposed by the available degrees of freedom, and which is at the expense of regional accuracy. Consequently, BOLD signals may be underestimated or possibly missed (Miller et al., 2005). Alignment of subcortical structures is substantially improved by a ROI-alignment (ROI-AL) approach, where segmentations of regions of interest (ROIs) are drawn on structural images and aligned directly, resulting in an increased statistical power (Yassa and Stark, 2009). The last iteration of this alignment tool is ROI-Demons, which has proven to be exceptionally accurate in the alignment of hippocampal subfields for instance (<http://darwin.bio.uci.edu/~cestark/roial/roial.html>). Thirion's original demons algorithm has been implemented by Vercauteren and enforces smooth deformations by operating on a diffeomorphic space of displacement fields (Thirion, 1998; Vercauteren et al., 2007). Here, we used the implementation of ROI-Demons in the DemonsRegistration command-line tool (<http://www.insight-journal.org/browse/publication/154>). Our segmented amygdalae ROIs were registered with a 'clean' participant (see ref (Prevost et al., 2011)) to serve as an initial model and to align all ROIs using DemonsRegistration. The resulting registered amygdalae were then averaged in SPM5 (using ImCalc) to create a first model (in this section, a model refers to an amygdalae template resulting from the averaged amygdalae ROIs). Subsequently, the initial non-registered masks were registered with this first model and the newly registered masks were averaged to create a second model. We repeated the last two steps three more times in order to generate a more accurate model. We finally registered our initial ROIs with the fifth model to generate the resulting displacement fields (or transformation calculations). These

individual displacement fields were then applied to each participant's normalized EPI scans in order to specifically normalize their amygdalae to our template mask. We applied the same transformation to each participant's structural scan before averaging all the aligned structural scans, to create an ROI-aligned average structural brain of our 26 participants. Finally, amygdalar subdivisions regions of interest (ROIs) were hand-drawn on our template amygdalae using the Atlas of the Human Brain (Mai et al., 2008). We delineated three sub-areas within the amygdala: the basolateral complex comprised of the basomedial, basolateral and lateral nuclei; the centromedial complex comprised of the central and medial nuclei; and the cortical complex (or cortical nucleus). In its most rostral part, the amygdala is exclusively composed of the basolateral complex. The cortical nucleus appears in the dorso-medial part of mid-rostral amygdala. The centromedial complex appears slightly more caudally than the cortical nucleus in the most dorsal part of the amygdala. The basolateral complex increases in size as one moves caudally from the anterior amygdala, has its maximal size midrostrocaudally and then decreases as one moves further back toward the caudal amygdala, whereas the cortical nucleus and centromedial complex slightly enlarge midrostrocaudally, but do not decrease in size as one moves further caudally within the amygdala. The cortical nucleus ends midcaudally, the basolateral complex ends in caudal amygdala while the centromedial complex ends in the most caudal part of amygdala.

3.2.8 fMRI data analysis

The event-related fMRI data were analysed by constructing sets of δ (stick) functions. All three phases of the task were modeled separately and here we report results only

from the PIT phase. The GLM included regressors at the time of trial onset and outcome delivery for both the instrumental and Pavlovian phases. For the PIT phase, we included regressors at the time of cue onset for 4 conditions: specific PIT when the option compatible with the Pavlovian cue elicited higher response rates, specific PIT when the incompatible option elicited higher response rates, general trials and neutral trials. All of these regressors were convolved with a canonical hemodynamic response function (HRF). The six scan-to-scan motion parameters derived from the affine part of the realignment procedure were included as regressors of no interest to account for residual motion effects. We also included thirteen additional regressors to account for physiological fluctuations (4 related to heart rate, 9 related to respiration) which were estimated using the RETROICOR algorithm (Glover et al., 2000). Sixteen of the 78 (3 sessions x 26 participants) log files could not be used to estimate these regressors due to a technical problem during data collection, and the missing physiological regressors were simply omitted for those sessions. Finally, to account for variance induced by differences in response rates across conditions in the instrumental and PIT phases, we added a regressor of no interest corresponding to the time course derived from this measure. This time course corresponded to the normalized number of response rates for each volume (period of 2s) of each session of each participant. All of these regressors were entered into a general linear model and fitted to each participant individually using SPM5. The resulting parameter estimates for regressors of interest were then entered into second-level one sample t-tests to generate random-effects level statistics. We then used an index of specific and general PIT for each participant as covariates at the second-level analysis on the contrast 'specific compatible > specific incompatible' for specific PIT and 'general trials > neutral trials' for general PIT. Separate covariates were used for specific and general PIT for each participant in a single 2nd level analysis. All reported fMRI statistics and p values arise from group

random-effects analyses. Two participants who did not respond to the incompatible option were excluded from this analysis. Note that among the 24 participants included, the minimum number of incompatible trials was 5. We present our statistical maps at a threshold of $p < 0.005$, corrected for multiple comparisons at $p < 0.05$. To correct for multiple comparisons in the striatum, we used small volume corrections (SVC) based on the coordinates reported by Bray et al. (2008) for specific PIT. In the amygdala, we first used the 3dFWHMx function in AFNI to estimate the intrinsic smoothness of our data, within the area defined by a mask corresponding to our amygdala template. We then used the AlphaSim function in AFNI to estimate via Monte Carlo simulation an extent threshold for statistical significance that was corrected for multiple comparisons at $p < 0.05$ for a height threshold of $p < 0.005$ within the amygdala ROI.

3.2.9 Plotting of parameter estimates

Plots of parameter estimates were extracted using the MarsBaR toolbox (<http://marsbar.sourceforge.net/>). For each participant, average parameter estimates were extracted within a 2mm sphere centered on the peak voxel of the amygdalar activity correlating with either specific or general PIT using the leave-one-participant-out method (Kriegeskorte et al., 2009), thereby avoiding a non-independence bias in the voxel selection in figure 3.4b and c, and without using this procedure in figure 3.4d and e. These beta estimates were correlated against the strength of behavioural specific and general PIT as illustrated in figure 3.4 b-e.

3.3 RESULTS

3.3.1 Behavioural results

3.3.1.1. Subjective ratings of food rewards

Each of the three food rewards was reported to be subjectively pleasant by the participants in ratings taken before starting the experiment, and after the test phase (two-tailed t-tests comparing the mean to zero, all $p < 0.001$, before: food 1: mean=1.75, std=1.46; food 2: mean=1.72, std=1.68; food 3: mean=1.49, std=1.24; after: food 1: mean=1.76, std=1.78; food 2: mean=1.85, std=1.61; food 3: mean=1.49, std=1.86). Furthermore, we performed a 2-way repeated measures ANOVA (2 time points x 3 different foods) to test whether there was a significant difference between the ratings for the different foods after as compared to before the experiment. We found no main effect for our factors time and food (all F 's < 0.7 , p 's > 0.5) as well as no interaction between these factors ($F(2,50)=0.09$, $p=0.9$), indicating that the foods remained equally subjectively pleasant for participants after the experiment. This was further confirmed by hunger ratings which showed that participants were not significantly more or less hungry after as compared to before starting the experiment (two-tailed paired t-test, $p=0.12$).

3.3.1.2 Instrumental conditioning

In the instrumental trials, both response rates and response times (i.e. the number of presses and the duration between trial onset and the first key press respectively) were not significantly different across the three response-outcome pairs (repeated-measures ANOVAs, response rate: $F(1,25)=1.31$, $p=0.26$; response time: $F(1,25)=0.62$, $p=0.44$), suggesting that participants were on average equally motivated to obtain each of the available outcomes. Note that the number of presses was normalized for each participant by subtracting the mean number of presses for this participant and dividing it

by the standard error. At the end of the experiment, participants had to indicate which instrumental response was associated with which food reward. On average, they were correct 91% of the time (std=14.64), which is significantly above chance (t-test comparing the mean to 50%, $p < 0.001$), suggesting that participants learned the specific action-outcome associations.

3.3.1.3 Pavlovian conditioning

Participants' heart rate was monitored using a pulse oximeter for the duration of the experiment. Existing research on heart rate responses to significant stimuli has identified an initial bradycardia associated with more aversive stimuli (Libby et al., 1973). This deceleration is thought to express attentional orienting to salient events through parasympathetic activity (Bradley, 2009). During the Pavlovian phase, non-rewarded trials were associated with a more pronounced cardiac deceleration (as assessed by the number of beats) compared to rewarded trials during anticipation, in a time window of 1.5-3.5s following stimulus onset, as reported elsewhere (Young et al., 2006) (two-tailed paired t-test, $p < 0.05$). Such physiological changes signal a difference in the elicited emotional state for cues presented on non-rewarded as compared to rewarded trials, thereby reflecting a differential conditioned response to the non-rewarded vs rewarded cues. Furthermore, at the end of the experiment, participants had to indicate which Pavlovian cue was associated with which food reward. On average, they were correct 93% of the time (std=20.28), which is significantly above chance (two-tailed t-test comparing the mean to 50%, $p < 0.001$), providing further evidence that participants learned the specific stimulus-outcome associations.

3.3.1.4 Pavlovian to Instrumental transfer effects

Each trial consisted of a 6 second duration where participants were performing a free-operant task, followed by a variable inter-trial interval of 2 to 6 seconds (**Fig. 3.1**). Behavioural evidence for specific PIT would be seen if the presence of the Pavlovian cue induced a higher response rate for the action associated with the same outcome as the Pavlovian cue. Conversely, evidence for general PIT would be seen if the presence of an appetitive Pavlovian cue induced a higher response rate for any action (despite being associated with a different outcome) than the presence of a cue that has not been paired with reward. This is exactly what we observed for both specific (two-tailed paired t-test, $p < 0.001$) and general (two-tailed paired t-test, $p < 0.05$) PIT (**Fig. 3.2a,b**).

Dividing the number of trials of the PIT phase into 5 blocks of 18 trials (as in Bray et al, 2008), we investigated whether the percentage in the number of presses significantly varied across these 5 blocks. A two-way repeated-measures ANOVA with the factors block and trial type was performed in order to investigate whether there was an effect of block during the extinction test. We found a main effect of trial type ($F(2,46)=99.21$, $p < 0.001$) but no main effect of block ($F(4,92)=0.58$, $p=0.68$), and no interaction between these two factors ($F(8,184)=1.13$, $p=0.36$), indicating that the biasing effect of the Pavlovian cues on response rates persisted for the duration of the extinction test and did not show any evidence of habituation (**Fig. 3.2c**). Furthermore, when investigating whether response rates for the three different keys differed in the instrumental and PIT phases using a 2-way repeated measures ANOVA (two phases x three response keys), we found no main effect of the key or the phase (all F 's < 0.5 , p 's > 0.3) and no interaction of these factors ($F(2,50)=1.8$, $p=0.18$). Similarly, when performing the same analysis using response times, there was no main effect of either factor (all F 's < 0.6 , p 's > 0.5) and no interaction ($F(2,50)=0.62$, $p=0.54$). This indicates that even though participants did not receive any reward in the PIT phase, they were equally motivated as when they

were getting food rewards in the instrumental phase (Table 3.2).

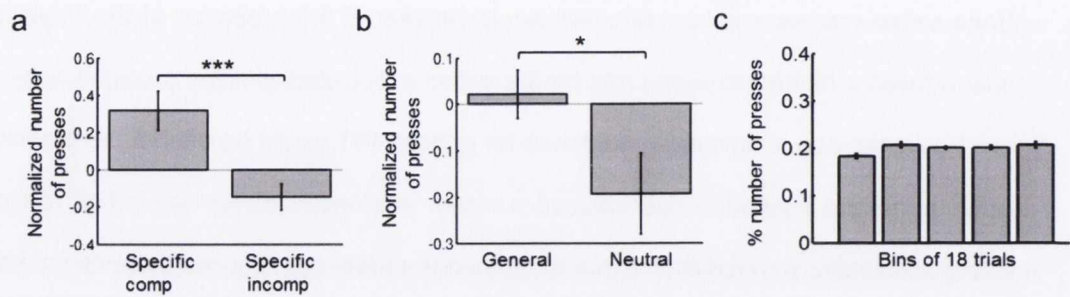


Figure 3.2. Behavioural results. *a*, Specific PIT. The response rate was significantly higher for the compatible responses than for the incompatible responses during specific trials. Error bars indicate SEM, *** indicates a significance of $p < 0.001$ as assessed by a two-tailed paired t-test. *b*, General PIT. The response rate was significantly higher for general trials than for neutral trials. Error bars indicate SEM, * indicates a significance of $p < 0.05$ as assessed by a two-tailed paired t-test. *c*, Percentage of number of presses for five 18-trial blocks across all trials. There is no significant linear trend across the session. Error bars indicate SEM.

	Mean RR during instrumental training	Mean RR during PIT phase
Participant 1	19.83 ± 1.46	15.18 ± 1.16
Participant 2	18.53 ± 1.58	13.16 ± 0.98
Participant 3	21.63 ± 1.36	18.32 ± 0.59
Participant 4	18.46 ± 0.81	15.44 ± 0.46
Participant 5	19.66 ± 1.05	22.05 ± 0.38
Participant 6	12.63 ± 0.66	18.17 ± 0.33
Participant 7	10.56 ± 0.67	22.17 ± 0.43
Participant 8	23.7 ± 0.58	22.06 ± 0.75
Participant 9	16 ± 0.60	16.88 ± 0.45
Participant 10	10.26 ± 0.62	15.92 ± 0.34
Participant 11	28.7 ± 0.49	27.68 ± 0.23
Participant 12	23.9 ± 1.88	26.4 ± 0.63
Participant 13	13.8 ± 0.68	23.03 ± 0.44
Participant 14	12.73 ± 0.68	12.63 ± 0.90
Participant 15	17.2 ± 1.03	23.62 ± 0.36
Participant 16	15.03 ± 0.71	21.12 ± 0.54
Participant 17	26.66 ± 1.28	24.96 ± 0.36
Participant 18	15.9 ± 0.52	22.56 ± 0.49
Participant 19	25.53 ± 1.36	26.12 ± 0.23
Participant 20	10.16 ± 0.57	9.32 ± 1.16
Participant 21	10.8 ± 0.78	18.68 ± 0.80
Participant 22	25.4 ± 0.62	22.58 ± 0.45
Participant 23	28.43 ± 1.21	30 ± 0.47
Participant 24	12.53 ± 0.77	17.91 ± 0.44
Participant 25	18.66 ± 0.96	10.56 ± 1.04
Participant 26	10.26 ± 0.52	3.33 ± 0.53

Table 3.2. Mean response rates (RR) ± SEM for each participant during the instrumental training and the PIT phase.

3.3.2 fMRI results

We report results from our analyses within the striatum and amygdala using a height threshold of $p < 0.005$, with an extent threshold significant at $p < 0.05$ corrected for multiple comparisons. We first examined the main effect of each type of PIT. To identify areas mediating specific PIT, we first compared activity when participants were

performing specific trials with activity when they were performing any other type of trial (i.e. general or neutral trials). As shown in figure 3.3, we found significant activity in an area of right ventrolateral putamen (MNI [x y z] [29 -2 -6], $T = 3.61$, $k = 20$ voxels) overlapping with that reported by Bray et al. (2008). Next, we compared activity when participants responded more to the instrumental action compatible with the Pavlovian cue to trials when they responded more to the action associated with a different incompatible outcome. To detect regions mediating general PIT, we compared activity between general and neutral CS trials, that is when participants were presented with a cue that had been paired with a reward outcome that was different from the reward outcomes associated with the available actions, compared to when the Pavlovian cue presented was associated with no outcome. These main effect contrasts did not reveal any significant activity in the striatum or amygdala at our statistical threshold.

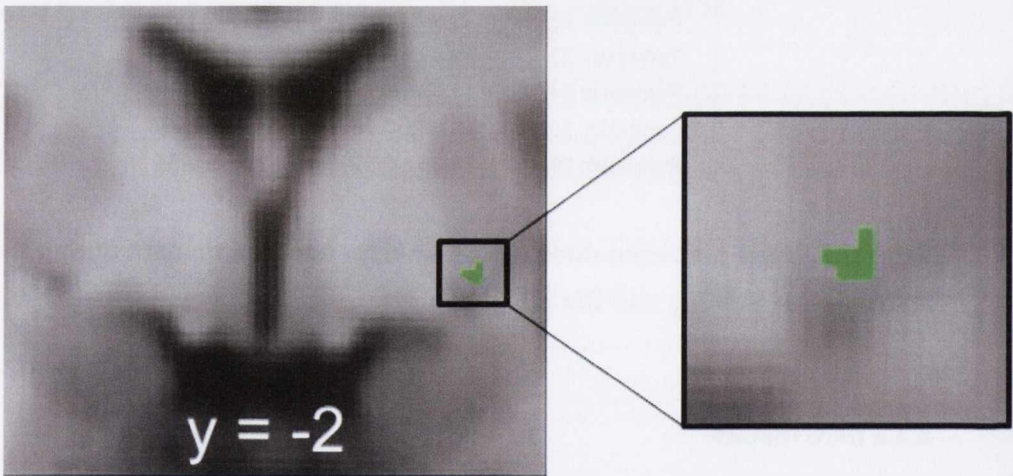


Figure 3.3. Blood oxygen level-dependent (BOLD) signal for the main effect of specific PIT ('specific trials > general & neutral trials') in right ventrolateral putamen, using a height threshold of $p < 0.005$, and a small volume correction significant at $p < 0.05$ corrected for multiple comparisons.

Similarly to the between-participant analysis carried out in Talmi et al. (2008), we next examined the extent to which activity within the amygdala in the relevant contrasts was modulated as a function of the variation across participants in the extent to which they exhibited behavioural evidence of general and specific transfer, by including our behavioural measures of specific vs general PIT for each participant as a covariate in these contrasts. Activity in a part of ventral amygdala consistent with the location of the basolateral complex (Mai et al., 2008) showed significant correlations with the magnitude of specific PIT exhibited by participants (**Fig. 3.4a**, in green, MNI [x y z] [-18 -3 -22], $T = 5.16$, $k = 24$ voxels). Conversely, a region of dorsal amygdala consistent with the location of the centromedial complex was found to be correlated with the degree of general PIT exhibited by participants (**Fig. 3.4a**, in red, MNI [x y z] [-15 -10 -11], $T = 5.83$, $k = 38$ voxels).

Finally, we tested for a differential involvement of BLA and CeN in specific and general PIT. When correlating the mean difference in the beta estimates from our general PIT contrast between our CeN ROI and our BLA ROI against participants' index of general PIT, we found a significant positive correlation ($r=0.48$, $p<0.05$), indicating that CeN is significantly more involved in general PIT than BLA. However, we did not find evidence suggesting that BLA is significantly more involved in specific PIT than CeN when correlating the mean difference in the beta estimates from our specific PIT contrast between our BLA ROI and our CeN ROI against participants' index of specific PIT ($r=0.24$, $p=0.26$).

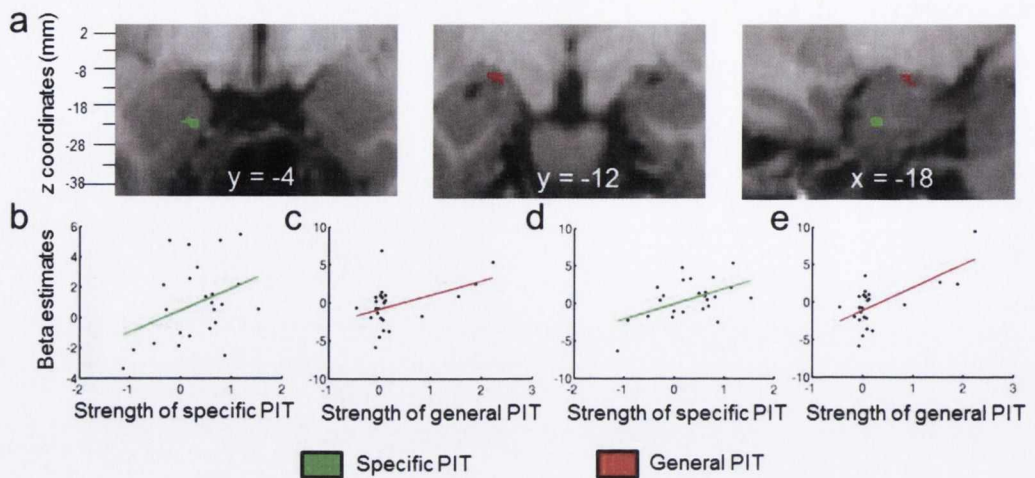


Figure 3.4. Specific and General PIT in the amygdala. *a*, BOLD signal positively correlating with the magnitude of specific (in green) and general (in red) PIT across participants in the ventral and dorsal amygdalae respectively, using a height threshold of $p < 0.005$, with an extent threshold significant at $p < 0.05$ corrected for multiple comparisons. *b*, Scatter plot showing the beta estimates extracted from a 2mm sphere centered on the peak voxel within the ventral amygdala correlating with specific PIT against the strength of behavioural specific PIT for each participant. *c*, Scatter plot showing the beta estimates extracted from a 2mm sphere centered on the peak voxel correlating with general PIT in the dorsal amygdala against the strength of behavioural general PIT for each participant. The plots in *c* and *d* were produced using a leave-one-participant-out procedure thereby avoiding non-independence issues (see Methods for details). *d,e* Similar scatter plots as displayed in *b,c* except that no leave-one-participant-out procedure was used.

3.4 DISCUSSION

In the current study, we show the first behavioural evidence of general appetitive Pavlovian instrumental transfer in humans, replicate findings of ventral striatum involvement in specific PIT (Bray et al., 2008) and further provide evidence for the involvement of amygdala in mediating specific and general PIT effects by making use of

a high-resolution imaging protocol. More specifically, we found a main effect of specific PIT in the ventrolateral putamen whereas we found between-participant differences for specific PIT in a region of human ventral amygdala within the boundaries of the BLA, and for general PIT in the dorsal amygdala within the boundaries of the CeN.

The ventral striatum has long been shown to be implicated in mediating PIT effects in rodents (Corbit and Balleine, 2011; Corbit et al., 2001; Hall et al., 2001) with recent evidence for a functional dissociation between the shell and core part of the nucleus accumbens in the mediation of specific and general PIT respectively (Corbit and Balleine, 2011). In humans, ventrolateral putamen activity for specific PIT has previously been reported (Bray et al., 2008), resonating with our findings.

Although we found a main effect of specific PIT in the ventral striatum, we did not find a main effect of specific or general PIT in the amygdala. Instead, we found an interaction between the BOLD signal in the amygdala and behaviour. It is interesting to note that Talmi et al. (2008) found a similar correlation between amygdala activity and the magnitude of their PIT effect, although the design of their study did not allow them to separately assess the effects of general and specific PIT. The results of these two studies are in agreement with the idea that amygdala involvement in PIT appears to depend on individual differences in behavioural expression of these effects, indicating that amygdala activity in PIT appears to be closely coupled to the degree of behavioural expression of these phenomena. Here, we expand on this earlier finding by showing that different types of PIT (specific versus general) seem to correlate with activity in different areas of the amygdala (BLA versus CeN). These results also dovetail with the results of a rodent lesion study in which it was found that lesions to the BLA impaired specific PIT while lesions of the CeN impaired general PIT (Corbit and Balleine, 2005). Therefore, the present findings indicate that a similar differentiation of function may be

present within the human amygdala, suggesting that the neural underpinnings of the influence of Pavlovian cues on decision making within the amygdalae are highly preserved across mammalian species.

These findings are also important for furthering understanding of the different functional contributions of amygdalar subregions in incentive processing in both humans and other animals. Killcross and Balleine have proposed a model of amygdala function on the basis of extant rodent lesion data in which the BLA is suggested to be more involved in mediating processing of specific sensory features of an outcome, while the CeN is proposed to be more involved in mediating processing of the general affective value of an outcome in a manner analogous to the distinction between consummatory and preparatory conditioning first espoused by Konorski (Konorski, 1967). In preparatory conditioning, a conditioned cue elicits generalized conditioned responses to appetitive or aversive outcomes such as approach and avoidance, while in consummatory conditioning, conditioned responses which are specific to the particular outcome predicted are elicited. Specific and general PIT effects have further been suggested to operate via these distinct incentive-processing mechanisms. Whereas in specific PIT, a conditioned stimulus is suggested to elicit sensory specific features of an outcome which in turn biases instrumental actions toward that outcome, in general PIT, the conditioned stimulus instead is suggested to elicit a generalized representation of the incentive value of that outcome in a manner independent of outcome identity (Balleine and Killcross, 2006; Dickinson and Balleine, 2002; Holland, 2004). Thus, the present findings are broadly consistent with the proposed differential contribution of the BLA and CeN in specific vs general incentive processing as proposed in the Killcross and Balleine model (Balleine and Killcross, 2006).

An important issue to note in the present study (and indeed any study involved

in measuring specific and general PIT) is that because these phenomena manifest themselves as very different behaviours, such that one is manifested through an overall increase in response rates while the other is measured in the current study by assessing bias in choice behaviour, these two phenomena may be scaled differently and therefore a direct contrast between the slopes of the degree of expression of specific versus general PIT is not meaningful. As a consequence, the appropriate test is to compare the slopes of activity in particular ROIs with respect to each of these phenomena separately (as opposed to directly comparing the slopes for general and specific PIT). When testing for a significant difference in the slopes of general PIT between the two regions, we found that CeN was significantly more involved in general PIT than was the BLA. In a separate analysis, we tested whether our BLA ROI was significantly more correlated with specific PIT than the CeN was correlated with specific PIT, and did not find statistical support for this hypothesis. Accordingly, we must acknowledge the limitation that we can only claim a partial but not a *double* dissociation between CeN and BLA with regard to specific versus general PIT. A likely explanation for the absence of a double dissociation is that unlike the CeN ROI which is circumscribed due to its small spatial size, the BLA ROI lacks sensitivity because it encompasses a considerable territorial expanse of the amygdala and therefore includes many voxels that do not have any involvement with specific PIT. Thus it is likely that specific PIT effects are being drowned out in the average across the BLA ROI.

This caveat notwithstanding, our results still enable us to make the claim that 1) the human BLA is involved in specific PIT while 2) the human CeN is involved in general PIT, and furthermore that the CeN is significantly more involved in general than specific PIT. Our findings have important implications for addictive and economic behaviour. Indeed, continued drug-seeking behaviour, particularly in the presence of environmental drug-associated cues, has been considered to reflect PIT influence

(O'Brien et al., 1998), and so has the impact of marketing strategies on consumer actions (Smeets and Barnes-Holmes, 2003). More generally, these results show how hr-fMRI can, when combined with appropriate behavioural paradigms, be deployed in order to address circuit-level questions about the functions of subregions within heterogeneous subcortical structures.

CHAPTER 4 DIFFERENTIAL INVOLVEMENT OF AMYGDALAR SUBREGIONS IN CONTEXT-LEARNING: EVIDENCE FROM HIGH-RESOLUTION IMAGING

Learning about whether one is in a positive or negative context is critical to survival. The animal literature suggests that the amygdala may be a strong candidate in encoding context-learning, with some evidence for a differential contribution of the basolateral (BLA) and central (CeN) areas in the encoding of appetitive and aversive contexts respectively. However, the role of the human amygdala in context-learning has remained elusive and the function of amygdalar subregions has been largely unexplored due to limitations in the spatial resolution of fMRI techniques. Here, we adopted a high-resolution fMRI (hr-fMRI) approach to investigate the functional role of distinct human amygdalar subregions in the encoding of context-learning, using a simple Pavlovian conditioning paradigm. We provide evidence for a differential involvement of the BLA and CeN complexes of the amygdala in context-learning, with the former mediating learning about positive context and the latter mediating learning about negative context. These findings shed considerable new light on the functions of human amygdalar subregions.

4.1 INTRODUCTION

Learning about the significance of stimuli in the environment has a clear adaptive function, probably even more so when those stimuli are predictive of a positive or negative context which may engender important consequences for animal survival. A

brain region that has received particular attention in the encoding of emotionally relevant contexts in both rodent lesion studies (Holland et al., 2002; Petrovich et al., 2009; Phillips and LeDoux, 1992) and neuroimaging studies in humans (Dolcos et al., 2004; Erk et al., 2003; Hardin et al., 2009; Kensinger and Schacter, 2006) is the amygdala. More specifically, recent rodent lesion studies have indicated that while the basolateral area (BLA) of the amygdala is involved in forming associations between Pavlovian cues and positive outcomes, the central nucleus (CeN) is involved in forming associations between Pavlovian cues and negative outcomes (Holland et al., 2002; Petrovich et al., 2009). However, an understanding of the role of distinct amygdalar subregions in contextual learning has remained lacking in humans due to limitations in the spatial resolution of fMRI techniques and whether learning about differently valenced contexts involves distinct human amygdalar subregions has yet to be explored.

The aim of the present study was to assess contextual learning in human amygdalar subregions using a high-resolution fMRI (hr-fMRI) protocol specifically optimized for the amygdala. We scanned human participants with fMRI while they participated in a Pavlovian conditioning procedure which involved learning about an appetitive or aversive context through associations between different stimuli and pleasant, neutral or aversive juice outcomes. In this task, participants were presented with cues leading to 1) the probabilistic delivery of affectively pleasant or affectively neutral liquid flavor outcomes or 2) to affectively unpleasant or affectively neutral outcomes in an appetitive and an aversive conditioning procedure respectively. Each session began with the learning of a specific context where two novel cues associated with either the affectively significant flavor liquid or the neutral liquid were presented. These associations reversed after an unpredictable number of trials. After another unpredictable number of

trials following reversal onset, a novel pair of cues was introduced, whose associations reversed following the same rule. Likewise, a third novel pair of cues was introduced wherein associations with the flavored liquids also reversed once. Every time a new pair of cue was introduced, participants had to learn about whether they were in an appetitive or an aversive context. To measure contextual learning in the amygdala, we tested for a decreasing response profile following cue onset every time a new pair of cue was introduced. To do so, we used exponential functions because they have successfully captured the initial learning of stimuli significance in the past, especially in fear conditioning studies in humans (Buchel et al., 1998; Morris et al., 2001a; Morris and Dolan, 2004).

We hypothesized that BLA would be preferentially involved when learning about the significance of a cue in a positive context (i.e. leading to an appetitive outcome) whereas the CeN would be preferentially involved when learning about the significance of a cue in a negative context (i.e. leading to an aversive outcome).

4.2 MATERIAL AND METHODS

4.2.1 Subjects

Nineteen right-handed subjects with a mean age of 22.21 ± 3.47 participated in the study. All subjects were free of neurological or psychiatric disorders and had normal or correct-to-normal vision. Written informed consent was obtained from all subjects, and the study was approved by the Trinity College School of Psychology Research Ethics committee.

4.2.2 Task description

Subjects participated in a Pavlovian task where they had to learn associations between different cues (fractal images) and a pleasant (Ribena), affectively neutral (artificial saliva made of 25mM KCl and 2.5 mM NaHCO₃) or unpleasant (salty tea made of 2 tea bags and 29g of salt per liter) flavor liquid. The task consisted of two sessions lasting approximately 22 minutes each. Each session was composed of 120 trials, leading to a total of 240 trials. In one of the sessions, subjects underwent an appetitive Pavlovian conditioning procedure whereby they were presented with cues leading to the subsequent delivery of either the pleasant flavor, or the affectively neutral one. In the aversive conditioning session, subjects underwent an aversive conditioning procedure whereby they were presented with cues leading to the subsequent delivery of either the unpleasant flavor stimulus, or else the affectively neutral stimulus. The rationale for including the appetitive and aversive conditioning procedures in separate sessions as opposed to including both conditions intermixed within the same sessions was to avoid contrast effects observed in prior behavioural piloting whereby cues signalling the aversive outcome tended to overwhelm cues signalling the pleasant one such that both the pleasant and the neutral cue stimuli were viewed as relief stimuli (contrasted against the aversive outcome) (Seymour et al., 2005). Performing the appetitive and aversive conditioning procedures in separate sessions ensured robust behavioural conditioning in both the appetitive and aversive cases and largely avoided contrast effects between the appetitive and aversive conditions.

For both sessions, on each trial, a cue was displayed randomly on either the left or right side of a fixation cross for 4 seconds. Following a well-established Pavlovian conditioning protocol (Gottfried et al., 2002; O'Doherty et al., 2004), subjects were also instructed to indicate on which side of the screen a given cue was presented by means

of pressing either a left or right button on a response box, yet they were also instructed that the subsequent outcomes were not contingent on their responses. This allows one to monitor the extent to which participants are paying attention to the cues on each trial. The offset of the cue (after 4 secs) was followed by delivery of one of the liquid flavor stimuli with a probability of 0.6, or else no liquid stimulus was delivered. The next trial was triggered following a variable 2-11 secs inter-trial interval.

At the beginning of each session, subjects were presented with two novel fractal cues (not seen before in the course of the experiment): which we will denote as cue 1 and cue 2. In the Appetitive session, cue 1 predicted the subsequent presentation of the pleasant liquid 60% of the time (or no liquid delivery 40% of the time), while in the Aversive session cue 1 predicted the delivery of the aversive liquid 60% of the time or else no liquid delivery 40% of the time. Cue 1 and Cue 2 trials were presented in a randomly intermixed order. After 16 trials (8 trials of each type), a reversal of the cue-outcome associations was set to occur with a probability of 0.25 on each subsequent trial. The probabilistic triggering of the reversal after the 16th trial ensured that the onset of the reversal was not fully predictable by subjects. Once a reversal was triggered, cue 1 no longer predicted the appetitive or aversive outcome but instead was associated with delivery of the neutral outcome, while cue 2 now predicted the appetitive or aversive outcome. After another 16 trials following the onset of the reversal, another event was triggered to occur with probability 0.25 on one of the subsequent trials: this time instead of a reversal, a completely novel pair of stimuli was introduced. Cue 3 was now paired with the appetitive or aversive outcome, while cue 4 was now paired with the neutral outcome. These new cues were presented for a further 16 trials, and followed again after a probabilistic trigger of $p=0.25$ on each subsequent trial with a reversal of the associations. After the reversal, a new set of cues were introduced

according to the same probabilistic rule and this was followed again by a reversal. Thus in total, 3 unique pairs of stimuli were used in each session and each of these pairs underwent a single reversal (**Figure 4.1**). A completely different set of cues were used for each session, so that subjects experienced a total of 6 pairs of fractal stimuli throughout the whole experiment. The order of the sessions was counterbalanced across subjects so that half of the subjects started the experiment with the appetitive session and half of the subjects with the aversive session.

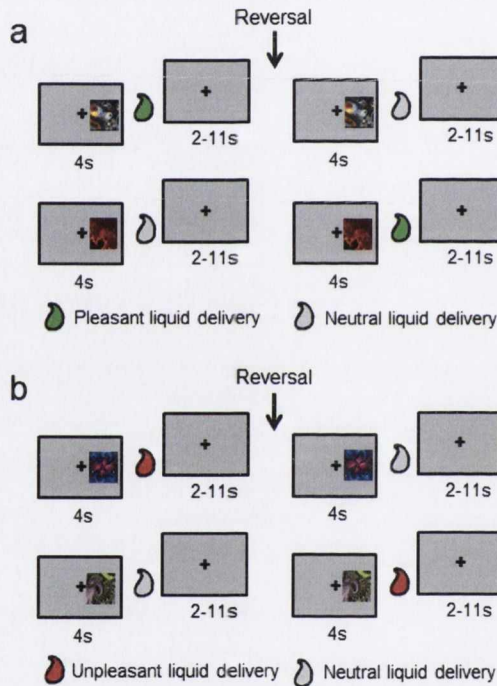


Figure 4.1. Appetitive versus aversive Pavlovian learning task. Sequence and timing of events in the appetitive (a) and aversive (b) sessions. On each trial, a cue was presented on one side of the screen for 4s, followed by some liquid delivery 60% of the time. The trial ended with a 2-11s inter-trial interval. Each session started with the presentation of cue 1 and cue 2, leading 60% of the time to a pleasant or a neutral liquid delivery in the appetitive session or an unpleasant or a neutral liquid delivery in the aversive session. After a number of trials, a reversal occurred so that cue 1 now led to the liquid associated with cue 2, and cue 2 led to the liquid associated with cue 1.

Subsequently, a new pair of cues was presented, which also reversed after a number of trials. In total, three new pair of cues were presented, and each of these pairs reversed once.

4.2.3 Apparatus

The pleasant, neutral and unpleasant tasting liquids were delivered by means of three separate electronic syringe pumps positioned in the scanner control room. These pumps pushed 1 mL of liquid to the subject's mouth via ~10 m long polyethylene plastic tubes, the other end of which were held between the subject's lips like a straw, while they lay supine in the scanner.

4.2.4 Fluctuations in heart rate

Estimates of heart rate were recorded using a pulse oximeter positioned on the forefinger of subjects' left hand. The time courses derived from this measure were used as a physiological index of conditioning (Nicola et al, 2006) as well as being used to remove physiological noise from the fMRI data analysis (see fMRI data analysis).

4.2.5 Data Acquisition

Functional imaging was performed on a 3T Philips scanner equipped with an 8-channel SENSE (sensitivity encoding) head coil. Since the focus of our study was on the amygdala, we only acquired partial T2*-weighted images centred to include the amygdala while subjects were performing the task. Nineteen contiguous sequential ascending slices of echo-planar T2*-weighted images were acquired in each volume,

with a slice thickness of 2.2 mm and a 0.3 mm gap between slices (in-plane resolution: 1.58 x 1.63 mm; repetition time (TR): 2000 ms; echo time (TE): 30 ms; field of view: 196 x 196 x 47.2 mm; matrix: 128 x 128). A whole-brain high-resolution T1-weighted structural scan (voxel size: 0.9 x 0.9 x 0.9 mm) and three whole-brain T2*-weighted images were also acquired for each subject. To address the problem of spatial EPI distortions which are particularly prominent in the medial temporal lobe (MTL) and especially in the amygdala, we also acquired gradient field maps. To provide a measure of swallowing motion, a motion-sensitive inductive coil was attached to the subjects' throat using a Velcro strap. The time course derived from this measure was used as a regressor of no interest in the fMRI data analysis. Finally, to account for the effects of physiological noise in the fMRI data, subjects' cardiac and respiratory signals were recorded with a pulse oximeter and a pressure sensor and further removed from time-series images. We discarded the first 3 volumes before data processing and statistical analysis to compensate for the T1 saturation effects.

4.2.6 Preprocessing

See section 3.2.7.

4.2.7 fMRI data analysis

The event-related fMRI data were analysed by constructing sets of δ (stick) functions at the time of cue presentation and at the time of outcome. In order to capture amygdala activity following the initial learning of a new context separately for the appetitive and aversive sessions, we constructed additional regressors by using

exponential functions as modulating parameters at the time of cue presentation every time a new pair of cues was presented. We used exponential functions because they have provided evidence for the decreasing response profile of amygdala following initial learning in humans numerous times, most notably in the aversive context of fear conditioning (Buchel et al., 1998; Morris et al., 2001a; Morris and Dolan, 2004). To demonstrate that amygdala activity accounts for context-learning as opposed to mere learning of new stimulus-outcome associations, we used similar functions following reversals of the stimulus-outcome associations as our second modulating parameter. All of these regressors were convolved with a canonical hemodynamic response function (HRF). The six scan-to-scan motion parameters derived from the affine part of the realignment procedure were included as regressors of no interest to account for residual motion effects. To account for motion of the subjects' throat during swallowing, we added a regressor of no interest for swallowing motion. Finally, we also included thirteen additional regressors to account for physiological fluctuations (4 related to heart rate, 9 related to respiration) which were estimated using the RETROICOR algorithm (Glover et al., 2000). Six of the 38 (2 sessions x 19 subjects) log files could not be used to estimate these regressors due to a technical problem during data collection, and the missing physiological regressors were simply omitted for those sessions. All of these regressors were entered into a general linear model and fitted to each subject individually using SPM5. The resulting parameter estimates for regressors of interest were then entered into second-level one sample t-tests to generate the random-effects level statistics used to obtain the results shown in **figure 4.2**. All reported fMRI statistics and *p* values arise from group random-effects analyses.

4.2.8 ROI analyses

Beta estimates from the basolateral complex and centromedial complex were estimated for each subject for both the appetitive and aversive sessions. These were then averaged across subjects. The repeated-measures ANOVA revealing the interaction subregion x context was performed in SPSS 16.0.

4.3 RESULTS

4.3.1 Behavioural evidence of Pavlovian conditioning

Participants' heart rate was monitored using a pulse oximeter for the duration of the experiment. Existing research on heart rate responses to significant stimuli has identified an initial bradycardia associated with more aversive stimuli (Libby et al., 1973). This deceleration is thought to express attentional orienting to salient events through parasympathetic activity (Bradley, 2009). Aversive trials were associated with a more pronounced cardiac deceleration (as assessed by the number of beats) compared to appetitive trials during anticipation, in a time window of 1.5-3.5s following stimulus onset, as reported elsewhere (Nicotra et al., 2006) (paired t-test, $p < 0.01$). Such physiological changes signal a more aversive emotional state for aversive as compared to appetitive trials, thereby reflecting a conditioned response to the aversive cues.

4.3.2 fMRI results

Our results are reported using a height threshold of $p < 0.005$, with an extent threshold significant at $p < 0.05$ corrected for multiple comparisons.

First, we investigated context-learning by looking for evidence of a decrease in amygdala activity following the presentation of new cues indicative of a new context in

the appetitive and aversive sessions separately using exponential decays (as in (Buchel et al., 1998; Morris et al., 2001a; Morris and Dolan, 2004). We tested for regions within the amygdala positively correlating with this metric. While we found activity in an area within the right basolateral complex in the appetitive context (**figure 4.2a** in green, MNI [x y z] [26 -5 -18], $T = 4.09$), we found spatially distinct dorsal regions of the bilateral amygdala in the aversive context (**figure 4.2a** in red, [x y z] [16 -4 -15], $T = 3.84$; [x y z] [-23 -6 -14], $T = 3.78$). We next performed an independent ROI analysis using our amygdala template, subdivided according to the approximate anatomical locations of the BLA and CeN as defined in the Mai atlas (Mai et al., 2008). We tested for a subregion (BLA vs CeN) x context (appetitive vs aversive) interaction on the averaged parameter estimates for our context-learning signal within the designated amygdalar ROIs. This analysis yielded a significant subregion x context interaction (repeated-measures ANOVA, $F(1,18)=4.67$, $p<0.05$) such that BLA showed greater correlation with new learning in the appetitive context while CeN showed a significantly greater correlation with new learning in the aversive context.

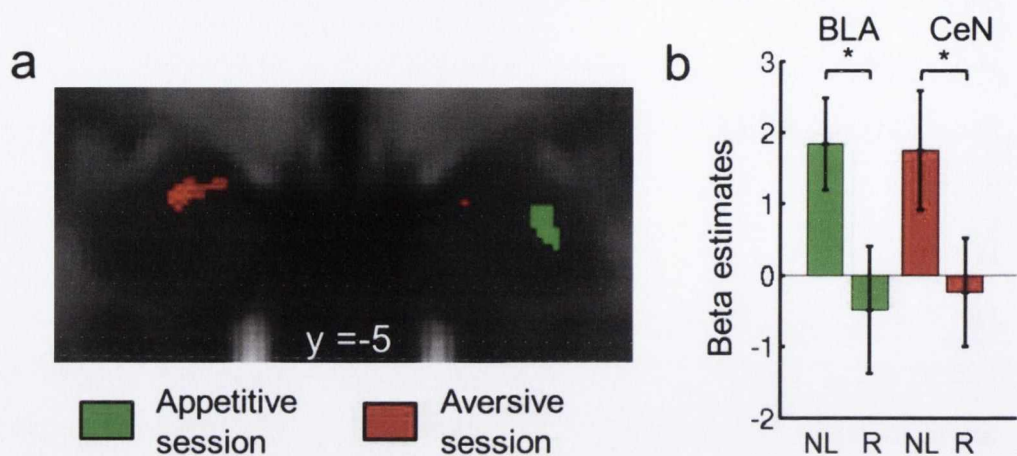


Figure 4.2. Amygdala signals following new learning. *a*, BOLD signals correlating with new learning (at the time of cue onset following the presentation of a new pair of cues) were found in BLA in the appetitive context (in green) and in CeN in the aversive context (in red). *b*, Plots showing the beta estimates following new learning (NL) and reversals (R) in the BLA ROI for the appetitive session (in green) and in the CeN ROI for the aversive session (in red). * indicates a significance of $p < 0.05$ as computed by one-tailed paired t-tests. Error bars represent the SEM.

Analogous exponential decays were used in our analysis following reversals. No area showed a significant correlation with these signals in the amygdala in both the appetitive and the aversive contexts at our statistical threshold. **Figure 4.2b** shows that BLA activity was significantly higher following new learning than reversals in the appetitive context whereas CeN activity was significantly higher following new learning than reversals in the aversive context (one-tailed paired t-tests, $p < 0.05$).

4.4 DISCUSSION

In recent years, it has become clear that the amygdala is involved in the initial learning of the significance of emotionally-relevant stimuli (Kensinger and Schacter, 2006; LeDoux, 2007; Schoenbaum and Roesch, 2005). However, systematic comparisons between this initial learning in appetitive versus aversive contexts on the one hand and between amygdalar subregions on the other hand have remained lacking in humans. Here, using an appetitive versus aversive Pavlovian task in combination with an hr-fMRI protocol specifically optimized for the amygdala, we provide evidence that the human amygdala is involved in contextual learning and further show that while BLA is implicated in learning about appetitive context, CeN is implicated in learning about aversive context.

Although our finding is not consistent with the serial model proposed by LeDoux (LeDoux, 2000; 2007), it suggests that different components of the amygdala may contribute differentially to learning about different contexts rather than being engaged in a serial manner. This is broadly consistent with recent rodent lesion data showing a double dissociation between BLA and CeN in mediating the influence of Pavlovian cues on the control of food consumption in an appetitive and aversive context respectively (Holland et al., 2002; Petrovich et al., 2009). Note that even though CeN may be more involved in learning about aversive contexts while BLA might be more involved in learning about positive ones, these differential contributions are unlikely to be exclusive as BLA and CeN are also highly interconnected and work together as a system. Furthermore, as shown in Chapter 3, it is entirely likely that this is not the only dimension on which these amygdala nuclei might be differentially engaged.

Moreover, our finding of differential activation between amygdalar subregions accords well with a recent neuroimaging study from our group which adopted an hr-fMRI protocol and reported a similar dissociation when participants were first exposed to a cue indicative of an appetitive or aversive context (Prevost et al, 2011; see also Chapter 5). More specifically, the initial presentation of a stimulus at the beginning of each session elicited stronger activity in BLA when it was predictive of a reward condition (leading to monetary gain) whereas it elicited stronger activity in CeN when it was predictive of an avoidance condition (leading to monetary loss). It is interesting to note that in the current study as well as in the study presented in Chapter 5, similar exponential decay functions were used, suggesting that from a computational perspective, the learning signal observed here may relate to the 'estimation uncertainty' metric reported in Chapter 5. However, as discussed in section 4.4, this decreasing response profile may be accounted for by alternative explanations such as novelty, arousal or orienting (although if in the case of novelty and arousal, one would not expect different areas to be involved as a function of valence).

Our finding that BLA is required to learn about appetitive context whereas CeN is required to learn about aversive context may relate to the distinct connectivity patterns of these subregions. Remarkably, BLA is highly interconnected with the ventral striatum (Fudge et al., 2002; Russchen et al., 1985) and these two structures have been shown to interact through dopamine-dependent mechanisms mediating reward learning (Cador et al., 1989). Correspondingly, CeN has been shown to act on the hypothalamus, midbrain and brainstem to mediate defense responses such as fight or flight reactions to potentially harmful stimuli (Kalin, 2004; Price and Amaral, 1981).

Consistent with the lack of amygdala activity increase following reversals, single-cell recording experiments have shown that neurons within the BLA fired during the initial

learning of significant Pavlovian cues (Saddoris et al., 2005; Schoenbaum et al., 1998, 1999, 2000) but not following reversals of the associations between these cues and outcomes (Schoenbaum et al., 2000). Likewise, lesions of the basolateral nucleus have been shown to prevent initial learning (Pickens et al., 2003; Schoenbaum et al., 2003b) but not reversal learning (Pickens et al., 2003; Schoenbaum et al., 2003a; Stalnaker et al., 2007). However, these studies have been largely confined to the BLA and systematic comparisons of neural responses across amygdala subregions have yet to be reported. Similarly, rapid decrease in the response profile of amygdala activity following initial presentations of Pavlovian stimuli has repeatedly been reported in neuroimaging studies of fear conditioning in humans, although the lack of spatial resolution prevented its attribution to a specific subregion (Buchel and Dolan, 2000; Buchel et al., 1998; Davis et al., 2010; LaBar et al., 1998; Morris et al., 2001a; Schiller et al., 2008). Nevertheless, consistent with our findings, there is no evidence of amygdala activity increase following reversals (Morris and Dolan, 2004; Schiller et al., 2008).

Taken together, our results provide considerable new insight into the functional contribution of BLA and CeN in humans and demonstrate the importance of adopting an hr-fMRI approach in order to gain a better understanding of human amygdala function.

CHAPTER 5 DIFFERENTIABLE CONTRIBUTIONS OF HUMAN AMYGDALAR SUBREGIONS IN THE COMPUTATIONS UNDERLYING REWARD AND AVOIDANCE LEARNING

To understand how the human amygdala contributes to associative learning, it is necessary to differentiate the contributions of its subregions. However, major limitations in the techniques used for the acquisition and analysis of fMRI data have hitherto precluded segregation of function within the amygdala in humans. Here, we used high-resolution fMRI in combination with an ROI-based normalization method to functionally differentiate the contributions of distinct subregions within the human amygdala during two different types of instrumental conditioning: reward and avoidance learning. Through the application of a computational-model based analysis, we found evidence for a dissociation between the contributions of the basolateral and centromedial complexes in the representation of specific computational signals during learning, with the basolateral complex contributing more to reward learning, and the centromedial complex more to avoidance learning. These results provide unique insights into the computations being implemented within fine-grained amygdala circuits in the human brain.

5.1 INTRODUCTION

The amygdala is composed of distinct subregions, each having unique inputs and outputs and each playing distinct functional roles in associative learning and the motivational control of behaviour (Cardinal et al., 2002; LeDoux, 2007). In humans, while much is now known about the functions of the amygdala as a whole (Adolphs,

2008; Bechara et al., 2003; Buchel and Dolan, 2000; Phelps, 2006; Seymour and Dolan, 2008), the precise functional contribution of its subregions remain unknown due to limitations in spatial resolution and analysis methods.

Here we focused on two distinct forms of instrumental conditioning: 1) reward-learning, in which an animal learns to select actions to obtain rewards, and 2) avoidance-learning, where an animal learns to select actions to avoid aversive outcomes. We focused on the differential contribution of the basolateral and centromedial complexes to these forms of learning. These areas have been differentially implicated in a variety of motivational and learning phenomena (Corbit and Balleine, 2005; Holland and Gallagher, 2003; Holland et al., 2002; Killcross et al., 1997; Petrovich et al., 2009). Although there is evidence for a differential contribution of these areas in the control of food consumption by appetitive and aversive learned cues respectively (Holland et al., 2002; Petrovich et al., 2009), the specific contribution of these regions in instrumental reward and avoidance learning remains largely unexplored.

To address this question, we used a high-resolution fMRI imaging procedure of the amygdala, together with an amygdala-specific normalization approach to functionally resolve activity in distinct amygdalar subregions across subjects, without the need for extrinsic spatial smoothing. Previous fMRI studies have reported computational learning signals in amygdala such as expected outcomes and prediction errors during performance of similar tasks (Elliott et al., 2004; Hampton et al., 2007; Seymour et al., 2005; Yacubian et al., 2006). Consequently, we hypothesized that we would find these different signals to be present within the amygdala in the present study.

Further, we hypothesized the presence of an uncertainty signal pertaining to the model-derived estimate of expected outcomes. Several pieces of evidence indirectly

point toward a possible role for the amygdala in encoding such a signal. Firstly, the amygdala has long been known to be involved in influencing the allocation of attention and orienting to conditioned stimuli, which could be driven by estimation uncertainty (Davis and Whalen, 2001; Dayan et al., 2000; Gallagher et al., 1990). Secondly, amygdala activation has been found when making decisions under conditions of high ambiguity (and hence uncertainty) about outcomes (Hsu et al., 2005). Finally, amygdala neurons have been implicated in encoding “associability” which is in essence a form of uncertainty about estimated outcomes (Pearce and Hall, 1980; Roesch et al., 2010). Thus, we postulated that the amygdala would also be involved in encoding a representation of uncertainty in the estimation of learned outcomes during task performance.

Going beyond previous fMRI studies, our use of high-resolution fMRI allowed us establish the extent to which these signals are localized to different subregions of the amygdala, both as a function of the type of computational signal involved, and as a function of the type of learning involved (whether reward or avoidance learning).

5.2 METHODS

5.2.1 Subjects

Twenty right-handed subjects (11 females) with a mean age of 22.65 ± 3.29 participated in the study. All subjects were free of neurological or psychiatric disorders and had normal or correct-to-normal vision. Written informed consent was obtained from all subjects, and the study was approved by the Trinity College School of Psychology Ethics Committee. Subjects were informed before performing the task that they would receive what they earned or lost during the task, added to an initial amount of 25 Euros. On average, subjects earned 31.5 ± 4.38 Euros.

5.2.2 Task description

Subjects participated in a probabilistic reversal learning task, which was divided into three sessions lasting approximately 16.5 minutes each, with 2 to 3 minutes break between sessions. Each session was composed of 72 trials, leading to a total of 216 trials. The task was equally divided in three conditions: a reward condition, an avoidance condition and a neutral condition, so that each condition was presented 24 times per session. The presentation order of these conditions was randomized across sessions and subjects. To ensure that each of the three conditions was presented an equal number of times per session, each condition was presented twice every 6 trials. On each trial, a fractal stimulus indicating the condition was displayed on the screen for 1-6 seconds. Subjects were then presented with a two-armed bandit slot machine and were asked to select the left or right arm. One action (left or right) was designated the correct action in that choosing it led to a monetary reward (+50 cents) 80% of the time and no monetary reward (0 cent) 20% of the time in the reward condition or no monetary reward (0 cent) 80% of the time and a monetary loss (-50 cents) 20% of the time in the avoidance condition. Consequently, choosing the correct action led to accumulating monetary gain in the reward condition and avoiding cumulating monetary loss in the avoidance condition. Choosing the incorrect action led to no monetary reward (0 cent) 80% of the time and a monetary reward (+50 cents) 20% of the time in the reward condition or a monetary loss (-50 cents) 80% of the time and no monetary reward (0 cent) 20% of the time in the avoidance condition. There was no correct/incorrect action in the neutral condition since all actions led to no monetary gain (0 cent). Subjects had one second to make their choice and subsequently waited for 1-8 seconds while the slot machine was spinning before the outcome was displayed for 1 second. The next trial started after a 1-6 s inter-trial interval (**Figure 5.1**). If subjects

chose the correct action on four consecutive occasions in a given condition, the contingencies reversed with a probability of 0.4 on each successive trial. Once reversal occurred, subjects then needed to choose the new correct action on four consecutive occasions before reversal could occur again (with 0.4 probability). Subjects were informed that reversals occurred at random intervals throughout the experiment but were not informed of the precise details of how reversals were triggered by the computer (so as to avoid subjects using explicit strategies, such as counting the number of trials to reversal). The subject's task was to accumulate as much money as possible and thus keep track of which action was currently correct in a given condition and choose it until reversal occurs. Subjects used a button box to select their left or right action. If subjects did not make their choice in the second provided, the current trial was aborted and the same trial was started again, to ensure that subjects did not avoid the trials in the avoidance condition. The assignment of the fractal images as cues in all three conditions was randomized across sessions and subjects.

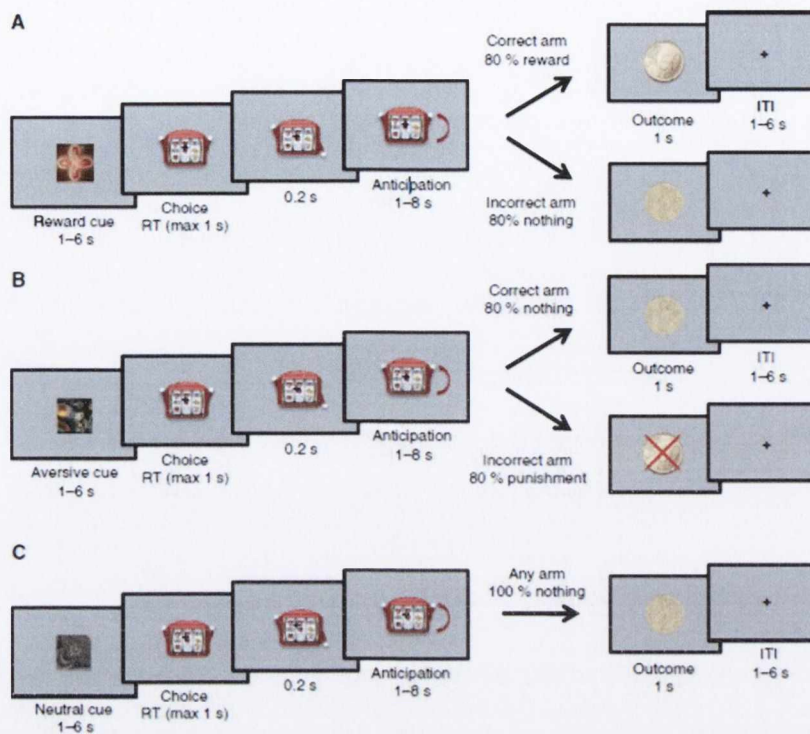


Figure 5.1. Reward/avoidance instrumental learning task. Sequence and timing of events of the Reward (**a**), Avoidance (**b**) and Neutral (**c**) conditions. The reward (**a**), aversive (**b**) or neutral (**c**) cue was shown for 1-6 s, followed by the choice of the left or right arm of the slot machine (maximum 1 s). The chosen arm went down for 200 ms after which the slot machine spun for 1-8 s. If the correct action was chosen, a 50 cents coin/a scramble coin (0 cent) was displayed for 1s 80% of the time in the Reward/Avoidance condition respectively whereas if the incorrect action was chosen, a scramble coin/crossed-out 50 cents coin (-50 cents) was displayed for 1s 80% of the time in the Reward/Avoidance condition respectively. All actions led to a scramble coin (0 cent) in the Neutral condition. The trial ended with a 1-6 s interval (ITI).

5.2.3 Prescan training

Before scanning, the subjects were trained on three different versions of the task. The first was a simple version of the reversal task, in which one of the two actions

yielded monetary rewards/no monetary reward 100% of the time, and the other action yielded no monetary reward/monetary losses 100% of the time in the reward/avoidance conditions respectively. These then reversed according to the same criteria as in the imaging experiment proper, in which a reversal was triggered with a probability of 0.4 after four consecutive choices of the correct action. This training phase was ended after the subject successfully completed two sequential reversals. The second training phase was identical to the first part except that no reversal occurred and that the outcome was probabilistic, as in the experiment. The training ended after the subject consecutively chose the correct action 8 times in a row. The final training phase consisted of the same task parameters as in the actual imaging experiment (stochastic outcomes, and stochastic reversals as described above). This phase ended after the subject successfully completed two sequential reversals. Different fractal stimuli were used in the training session from those used in the scanner. Subjects were informed that they would not receive remuneration for their performance during the training session.

5.2.4 Data Acquisition

Functional imaging was performed on a 3T Philips scanner equipped with an 8-channel SENSE (sensitivity encoding) head coil. Since the focus of our study was on the amygdala, we only acquired partial T2*-weighted images centered to include the amygdala while subjects were performing the task. These images also encompassed the ventral part of the prefrontal cortex, the ventral striatum, the insula, the hippocampus, the ventral part of the occipital lobe and the upper part of the cerebellum (amongst other regions). Nineteen contiguous sequential ascending slices of echo-planar T2*-weighted images were acquired in each volume, with an in-plane resolution

of 1.58 x 1.63mm, and a slice thickness of 2.2 mm and a 0.3 mm gap between slices; repetition time (TR): 2000 ms; echo time (TE): 30 ms; field of view: 196 x 196 x 47.2 mm; matrix: 128 x 128). A whole-brain high-resolution T1-weighted structural scan (voxel size: 0.9 x 0.9 x 0.9 mm) and three whole-brain T2*-weighted images were also acquired for each subject. To address the problem of spatial EPI distortions which are particularly prominent in the MTL and especially in the amygdala, we also acquired gradient field maps. To account for physiological fluctuations, subjects' cardiac and respiratory signals were recorded with a pulse oximeter and a pressure sensor placed on the umbilical region and further removed from time-series images (see fMRI data analysis below). We discarded the first 3 volumes before data processing and statistical analysis to compensate for the T1 saturation effects.

5.2.5 Preprocessing

See section 3.2.7.

Segmentation of Amygdalar subregions

To date, two different techniques have been proposed to anatomically delineate human amygdalar subregions using MRI: (Amunts et al., 2005) used post-mortem cytoarchitectonics to develop a probabilistic map of amygdala nuclei and other medial temporal lobe structures, while (Solano-Castiella et al., 2010) used diffusion-weighted MRIs to cluster amygdala subregions based on the predominant diffusion direction. However, although both of these approaches are promising future directions, in our case neither of these segmentation methods were deemed appropriate. The Amunts amygdala map was found to extend far beyond the limits of the amygdala into the white matter of the medial temporal lobes, suggesting a mismatch between the localization of

the amygdala in our sample and that in the probabilistic map. This may be due to the fact that the post-mortem brains used to derive the map were from a much older population (average age 65), than that used in our study (average age 23 years). Given that gross structural changes are known to occur in the brain with aging such as atrophication of brain structures and increase in ventricular size (Anderton, 1997), differences in amygdala shape and location between a younger sample (standard for fMRI studies on healthy controls) and an older sample might account for such effects, rendering the Amunts map unsuitable for our sample. The Solano-Castiella et al. diffusion-based segmentation method yielded two distinct areas of the amygdala along a medial to lateral dimension. Unfortunately this delineation does not map clearly onto the known nuclear structure of the amygdala and the known location of the basolateral and centromedial complexes (which are better differentiated along a ventral to dorsal dimension). This might reflect the possibility that the diffusion-based segmentation method is likely sensitive to the orientation of myelinated fibres as opposed to cytoarchitecture (Solano-Castiella et al., 2010). Because of these issues, we used a manual segmentation approach to delineate the amygdala sub-structures as follows:

Amygdalar subdivisions were hand-drawn on our template amygdalae using the Atlas of the Human Brain (Mai et al., 2008). We delineated three sub-areas within the amygdala: the basolateral complex comprised of the basomedial, basolateral and lateral nuclei; the centromedial complex comprised of the central and medial nuclei; and the cortical complex (or cortical nucleus). In its most rostral part, the amygdala is exclusively composed of the basolateral complex. The cortical nucleus appears in the dorso-medial part of mid-rostral amygdala. The centromedial complex appears slightly more caudally than the cortical nucleus in the most dorsal part of the amygdala. The basolateral complex increases in size as one moves caudally from the anterior

amygdala, has its maximal size midrostrocaudally and then decreases as one moves further back toward the caudal amygdala, whereas the cortical nucleus and centromedial complex slightly enlarge midrostrocaudally, but do not decrease in size as one moves further caudally within the amygdala. The cortical nucleus ends midcaudally, the basolateral complex ends in caudal amygdala while the centromedial complex ends in the most caudal part of amygdala.

In order to address the concern that our manual segmentation method could be prone to experimenter bias, we also performed a much more basic differentiation based on the very fundamental observation that the ventro-dorsal extent of the centromedial complex extends approximately across the top fourth of the volume of the whole amygdala, while the basolateral complex extends across the bottom three fourths of the volume. We therefore delineated much simpler ROIs whereby the top $\frac{1}{4}$ of the amygdala depicts the approximate location of the centromedial complex while the bottom $\frac{3}{4}$ represents the approximate location of the basolateral complex. Even when using these simpler ROIs, all of the condition $\times$ subregion interactions reported in the results in the ROI comparisons were still significant at the $p < 0.05$ level. This demonstrates that our results are extremely unlikely to be due to experimenter bias in the manual segmentation of the amygdala ROIs as described above.

5.2.6 Computational model-based analysis

5.2.6.1 Action value learning.

In order to best characterise the nature of learning in this task, several variants of the Rescorla Wagner/delta learning rule were considered (Rescorla, 1972): Simple delta rule learning (Gluck and Bower, 1988) consisting of only updating the chosen

action, decay reinforcement (Erev and Roth, 1998) which updates the chosen action and decays unchosen actions to 0, and fictive reinforcement learning (Fudenberg and Levine, 1998; Myung and Busemeyer, 1992), which updates the unchosen action with what would have been received had the action been selected, a model that has previously proven successful in explaining behavioural and neural activity during reversal learning (Glascher et al., 2009; Hampton et al., 2007). The fictive model assumes that outcomes available on the two actions are perfectly anticorrelated because of the reversal structure; for example, if the selected option on an appetitive trial resulted in no reward, then the alternate option is assumed to have provided a reward. The action values $Q(t)$ for both the selected and unselected choices were updated according to:

$$Q_j(t + 1) = Q_j(t) + \alpha \cdot (R(t) - Q_j(t)) \quad (1)$$

For each of these three variants, when j is the chosen action, α is the learning rate, and $R(t)$ is the reward received on the current trial. If the 'good' outcome was obtained on the current trial, $R(t) = 1$, else $R(t) = 0$. When j is the unchosen action, $\alpha = R(t) = 0$ for the simple delta rule model (resulting in no change in expectation). For both the decay and fictive models, α is the learning rate and $R(t) = 0$ or 1 – reward received on the current trial, respectively. The new action value at trial $t + 1$ for the chosen action is based on the sum of the current action value and the prediction error between the reward obtained and the action value at time t , weighted by the learning rate.

5.2.6.2 Action selection computation

To choose which of the two actions to select, the model compares action values and probabilistically selects the action expected to give the most reward in the future. The probability $P_j(t)$ of choosing action j out of the entire set of actions at time t is given by the softmax rule (Sutton and Barto, 1998), a parameterized version of the Luce choice axiom (Luce, 1959):

$$P_j(t) = e^{Q_j(t)/\theta} / \sum_i e^{Q_i(t)/\theta} \quad (2)$$

The inverse temperature parameter θ controls the extent to which decisions are consistent. As θ approaches 0, the model predicts increasingly deterministic choice whereas larger values of θ indicate increasingly random selection.

Plots of subject choices versus model predictions in the reward and avoidance conditions are shown in Figure 5.2c,d.

-Model fitting and comparison methods

The values for three free parameters were estimated for each model variant, using `fminsearch`, a simplex (Nelder and Mead, 1965) hill descent algorithm available in Matlab. Learning rates α_{Rew} and α_{Av} were fit separately for the reward and avoidance conditions as well as the softmax temperature parameter θ . Data from the neutral condition were not fit. Values for $Q(0)$ were set to 0.5, reflecting the idea that individuals were updating the probabilities of obtaining the outcomes, and high values of Q indicate a greater expectation for obtaining the superior outcome. The data was pooled across subjects, as this method has proven to more reliably estimate true underlying parameter values with smaller numbers (e.g., <100) of trials per subject (Cohen et al., 2008a).

Parameter estimation was conducted using maximum likelihood estimation. Here, we obtained the values for the parameters that minimised the negative log likelihood (NLL) of the data given the model. The NLL from each model was compared to that given by a baseline model that assumed random selection on each trial. Hence, the comparison was between random choice and ordered choice. Bayesian information criterion (BIC) was used to compare the baseline model to the learning model variants. The BIC adds a penalty proportional to the number of additional free parameters to the NLL of each model, depending also on the number of degrees of freedom which, in this case, is the total number of reward and avoidance trials across all subjects (Schwarz, 1978). Note that if across all subjects each action is chosen with equal frequency – a very realistic scenario in a reversal learning task, as used here – the only hope the learning models have in outperforming the baseline model is to detect outcome-based changes in choice behaviour (i.e learning). As shown in table 5.1, the decay reinforcement learning model best fit the data. Consequently, this model variant was used in the subsequent fMRI data analysis.

5.2.6.3 State value learning

We also modelled learning of the value of the states indicated by the presentation of the visual cues at the time of trial onset. These cues indicated either a reward trial, an avoidance trial, or a neutral trial. The state values were learned directly using a delta learning rule similar to that used to learn the action values; the values of the reward and avoidance state cues were updated separately.

The formula for the cue value updating is:

$$V_s(t+1) = V_s(t) + \delta \cdot \sqrt{U(t)/U(1)} \cdot (R(t) - V_s(t)) \quad (3)$$

where $V_s(t)$ is the value of the cue on trial t .

However, as the learning rate for the state values could not be estimated directly from the choice data as was the case for the action values, we adopted a reinforcement learning based approximation of a Kalman filter to compute the trial by trial uncertainty $U(t)$ in the estimate of $V_s(t)$ for the cues and used this estimation uncertainty signal in order to dynamically adapt the learning rate parameter for the update of $V_s(t)$ in an approximately optimal manner. Estimation uncertainty captures the extent to which one's expectations about future reward vary over time. δ is a fixed learning rate equal to 0.2 and R is the reward (or outcome) obtained on each trial. The difference term represents a prediction error. The product term on the right side of the equation represents a weighted adjustment to the state value on the current trial, reflecting a dynamic adjustment of the impact of the prediction error, due to changes in the estimation uncertainty component. This state value updating was agnostic regarding the action selected. Estimation uncertainty was updated according to:

$$U(t + 1) = U(t) + \delta \cdot ((V_s(t + 1) - V_s(t))^2 - U(t)) \quad (4)$$

Note that estimation uncertainty is updated by a weighted difference of the squared deviation in state value expectations and the previous uncertainty value.

A similar procedure could have been used to generate learning rates for the action values as well as for the state values, however, for the action values, estimation of the learning rates directly from the behavioural data provided a better overall model-fit to the behavioural data than did the use of the estimation uncertainty approach ($BIC(\text{Decay RL model}) = 3751.6 < BIC(\text{estimation uncertainty model}) = 3805.8$, where a lower BIC value denotes superior performance). Hence, we deployed estimation uncertainty for computation of the learning rate parameter in the state value case only.

5.2.7 fMRI data analysis

The event-related fMRI data were analyzed by constructing sets of δ (stick) functions at the time of cue presentation, at the time of choice and at the time of outcome for the reward, avoidance and neutral conditions. Additional regressors were constructed by using the estimation uncertainty and estimated state values as modulating parameters at the time of cue presentation, the value for the chosen action as a modulating parameter at the time of choice and the prediction error as a modulating parameter at the time of outcome. All of these regressors were convolved with a canonical hemodynamic response function (HRF). The six scan-to-scan motion parameters derived from the affine part of the realignment procedure were included as regressors of no interest to account for residual motion effects. Finally, we included thirteen additional regressors to account for physiological fluctuations (4 related to heart rate, 9 related to respiration) which were estimated using the RETROICOR algorithm (Glover et al., 2000). Eight of the 60 (3 sessions x 20 subjects) log files could not be used to estimate these regressors due to a technical problem during data collection, and the missing physiological regressors were simply omitted for those sessions. All of these regressors were entered into a general linear model and fitted to each subject individually using SPM5. The resulting parameter estimates for regressors of interest were then entered into second-level one sample t-tests to generate the random-effects level statistics used to obtain the results shown in **Figures 5.3-5.5**. All reported fMRI statistics and p values arise from group random-effects analyses. We present all our statistical maps at a threshold of $p < 0.005$, corrected for multiple comparisons at $p < 0.05$. To correct for multiple comparisons, we first used the 3dFWHMx function in AFNI to estimate the intrinsic smoothness of our data, within the area defined by a mask corresponding to our amygdala template. We then used the AlphaSim function in AFNI

to estimate via Monte Carlo simulation an extent threshold for statistical significance that was corrected for multiple comparisons at $p < 0.05$ for a height threshold of $p < 0.005$ within the amygdala ROI. This extent threshold corresponds to $k = 26$ voxels.

In a separate analysis, we tested for responses in the amygdala to receipt of rewarding and punishing outcomes. We used exactly the same general linear model as mentioned above, except that we replaced the prediction error parametric modulators with outcome parametric modulators. In the reward condition, on any trial where a reward was obtained, this modulator was set to 1, whereas anytime a no-reward was obtained the modulator was set to -1. For the avoidance condition, any time no punishment was obtained, the regressor was set to 1, and when a punishment was obtained it was set to -1. Similar activation patterns were obtained in amygdala in response to outcomes (reward vs no-reward; no punishment versus punishment) as were obtained for prediction errors. This may be largely due to the fact that for this task, prediction errors and outcomes are intrinsically strongly correlated. However, in order to discriminate between them, we used a Bayesian model selection procedure (BMS), to test which signal (prediction error or outcome) better accounted for the observed activity. For the reward condition, we included voxels in this model comparison that were significantly correlated with *either* the outcome parametric regressors for the outcome model *or* with the reward prediction error signal in the prediction error model. This allowed us to input only those voxels responsive either to prediction errors or outcomes and to avoid a non-independence bias in the voxel selection. The same approach was adopted for the avoidance condition. Using the `spm_BMS` function in SPM8, we found that activity at the time of outcome was better explained by the prediction error signal at the time of outcome than by the receipt of outcomes themselves (Reward condition: exceedance probability = 0.739; Avoidance condition:

exceedance probability=0.742). Consequently, we interpret activity at the time of outcome as pertaining to prediction errors and not outcome values per se.

5.2.8 ROI analyses

Functional regions of interest (ROI) were defined using the MarsBaR toolbox (<http://marsbar.sourceforge.net/>). All ROIs are functional clusters of interest as they appeared on the statistical maps of a given contrast. Percent signal change was extracted for each subject from each of these ROIs and then averaged across subjects to plot action values (**Figure 5.3b**), prediction error (**Figure 5.4b**) and estimation uncertainty (**Figure 5.5c**) according to 6 categories (category 1 corresponding to the lowest values and category 6 corresponding to the highest values).

Beta estimates from different amygdalar subregions, namely the basolateral complex and centromedial complex, were estimated for each subject for both the reward and avoidance conditions. These were then averaged across subjects and beta estimates from the action value contrasts are plotted in **Figure 5.3c** and those from the estimation uncertainty contrasts are plotted in **Figure 5.5d**. Repeated-measures ANOVA showing the interactions subregion x condition were performed in SPSS 16.0. Note that because the volume of the basolateral and centromedial ROIs is substantially different (the basolateral is approximately 3 times larger than the centromedial ROI), when directly comparing the ROIs there is a possible violation of the sphericity assumption in the general linear model due to potential differences in the variance distributions. To account for this, we equalized the volume of the ROIs in an unbiased way by defining a sphere of the same volume of the centromedial ROI, placing this in the geometric centroid of the basolateral ROI, and then performing the comparison

between these two identical volumes. Nevertheless, even when taking the whole extent of the basolateral complex as the ROI, all the ROI effects reported in the results section remained significant at $p < 0.05$ both for the estimation uncertainty contrast and the action value contrast. Thus, all the results reported are robust to differences in the ROI volumes.

5.3 RESULTS

5.3.1 Behavioural results

Subjects showed behavioural evidence of learning in both the reward and avoidance conditions in that they favored choice of the reward maximizing or loss minimizing action over the alternative actions (**Figure 5.2a**). To further show that subjects were able to learn the task, we demonstrated that overall they performed better than would be expected by chance in both conditions (**Figure 5.2b**). We fit a variant of a reinforcement learning model (RL) to generate trial-by-trial estimates of the computations necessary to underpin performance on such a task, including the expected future reward corresponding to individual actions at the time of choice, and prediction errors signalling the difference between expected and actual outcomes used to facilitate updating of the estimated expected reward for each action. This model performed significantly better than a baseline model choosing purely randomly, even after adjusting for additional free parameters (Table 5.1). Figure 5.2**c,d** shows the RL model predictions versus participants' actual choices in the reward and avoidance conditions respectively. The trial-by-trial estimates of the action values for a typical subject are represented in figure 5.2**e**.

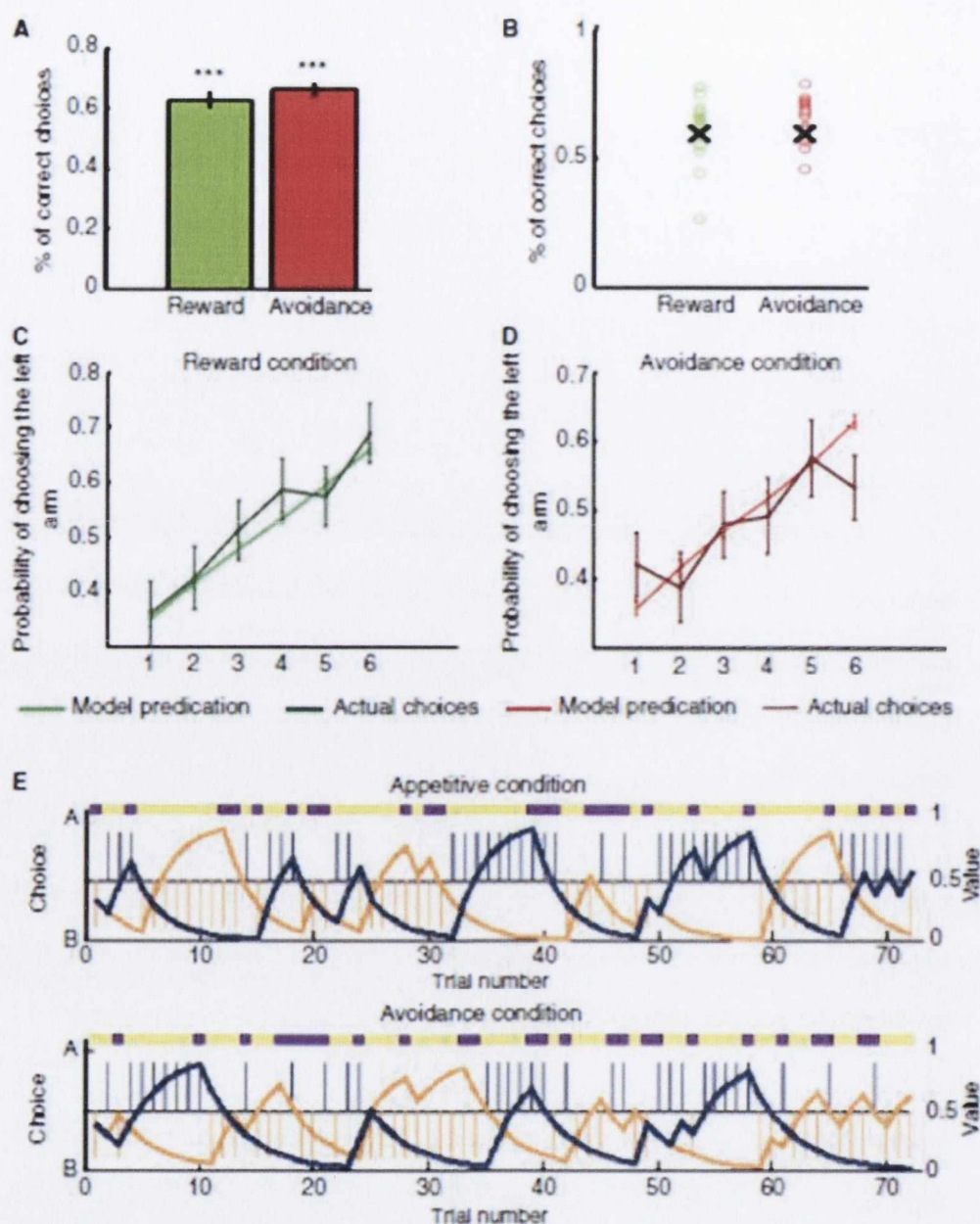


Figure 5.2. Behavioural results. *a*, Percentage of correct choices in the Reward and Avoidance conditions averaged across sessions and subjects. Error bars represent SEM. *** indicates a significance of $p < 0.001$ as computed by a one sample t-test comparing the mean of the correct choices for each condition against a mean of 0.5. *b*, Percentage of correct choices across sessions for each subject in the Reward and Avoidance conditions. Each subject is represented by a circle in each condition and the black cross indicates the threshold above

which subjects performed better than chance at $p < 0.05$. Note that the average percentage of correct choices does not exceed 70% because of the probabilistic nature of the task and the occurrence of reversals. **c,d**, Plots of subject choices (dark colors) versus decay RL model predictions (bright colors) in the Reward (**c**) and Avoidance (**d**) conditions for six categories (category 1 representing the lowest predictions and category 6 the highest predictions). Error bars represent SEM. **e**, Trial-by-trial behavioural model fit of the decay RL model for a typical subject. Actual choices of the different actions A (blue sticks) and B (orange sticks), and the model-predicted action value for each action (blue and orange lines for action A and B respectively) are shown. The colored bar on top shows whether subjects won 50 cents/avoided losing 50 cents in yellow or did not win 50 cents/lost 50 cents in purple in the Reward and Avoidance conditions respectively.

Model	α_{Rew}	α_{Av}	θ	BIC
Baseline	—	—	—	3992.5
Simple	0.75	1	0.06	3849.2
Decay	0.32	0.24	1.13	3751.6
Fictive	0.25	0.44	0.26	3800.6

Table 5.1. Model comparison between several variants of the delta learning model and a baseline model. Simple delta rule learning consists of only updating the chosen action, decay reinforcement updates the chosen action and decays unchosen actions to 0, and fictive reinforcement updates the unchosen action with what would have been received had the action been selected. A smaller BIC value indicates a better fit, therefore the decay reinforcement learning model best fit the data.

5.3.2 fMRI results

We found evidence for a number of distinct computational signals within the amygdala in the reward and avoidance conditions in our fMRI data. Our results are reported using a height threshold of $p < 0.005$, with an extent threshold significant at $p < 0.05$ corrected for multiple comparisons.

5.3.2.1 Action-value signals

We first examined BOLD activity in the amygdala at the time of choice. In the reward condition, an area on the ventral border of the right anterior amygdala was found to show significant correlations with the expected reward signal for the action chosen on each trial, according to the RL model (**Figure 5.3a** in green, MNI [x y z] [20 -1 -26], $T = 5.29$, $k=27$ voxels). A spatially distinct far dorsal region of the right amygdala was found to be positively correlated with the action value signal in the avoidance condition (**Figure 5.3a** in red, [x y z] [18 -2 -12], $T = 5.35$, $k=52$ voxels), such that the greater the activity in that area, the less an aversive outcome is predicted to occur. We next performed an independent ROI analysis using our amygdala template, subdivided according to the approximate anatomical locations of the basolateral complex, the centromedial complex and the cortical nucleus defined in the Mai atlas (Mai et al., 2008). We tested for a subregion (basolateral vs centromedial) x condition (reward vs avoidance) interaction on the averaged parameter estimates for our action value signal within the designated amygdalar ROIs. This analysis yielded a significant subregion x condition interaction ($F(1,19)=7.06$, $p<0.05$) such that the basolateral amygdala showed greater correlation with action value in the reward condition while the centromedial complex showed a significantly greater correlation with action value in the avoidance condition (**Figure 5.3c**). We also looked for areas correlating negatively with action value in both the reward and avoidance conditions, that is, areas showing an increase in activity the less reward was predicted to occur given the chosen action, according to the RL model. We found significant activity in a region of anterior dorso-medial amygdala in the possible vicinity of the cortical amygdala nucleus in the reward condition ([x y z] [17 3 -15], $T = 4.23$, $k=32$ voxels).

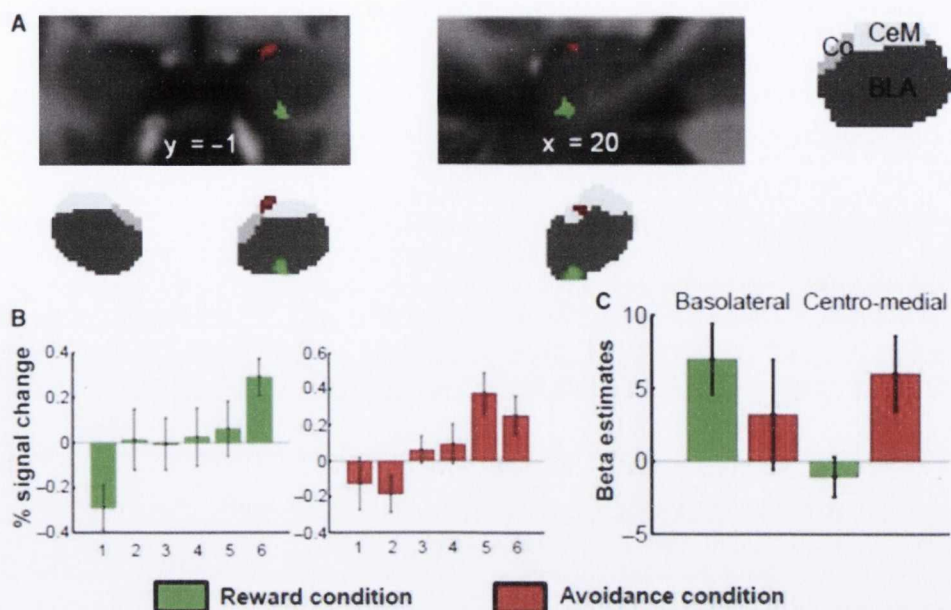


Figure 5.3. Action reward value signals. *a*, Blood oxygen level-dependent (BOLD) signals correlating with the magnitude of the expected reward value of the chosen action were found in the basolateral complex in the Reward condition (in green) and in the centromedial complex in the Avoidance condition (in red). *b*, Plots showing the percent signal change for 6 categories of action values (category 1 representing the lowest action values, and category 6 representing the highest action values) in the Reward (green) and Avoidance (red) conditions in the clusters activated. *c*, Beta estimates showing an interaction between the subregion (basolateral versus centromedial) and the condition (Reward versus Avoidance) ($p < 0.05$).

5.3.2.2 Reward prediction error signals

Next, we examined amygdala activity at the time of outcome. We tested for amygdala activity correlating with reward prediction errors during the reward and avoidance conditions. We found activity positively correlating with prediction error during both the reward and avoidance conditions within the basolateral complex ROI (reward condition: $[x \ y \ z] [-25 \ -4 \ -19]$, $T = 5.24$, $k=47$; $[x \ y \ z] [22 \ -8 \ -16]$, $T = 5.13$, $k=140$; $[x \ y \ z] [-21 \ -11 \ -19]$, $T = 4.50$, $k=96$; $[x \ y \ z] [28 \ -1 \ -19]$, $T = 4.27$, $k=47$; avoidance

condition: $[x\ y\ z]$ $[-25\ -5\ -23]$, $T = 5.01$, $k=38$; $[x\ y\ z]$ $[-15\ -6\ -21]$, $T = 4.88$, $k=39$; $[x\ y\ z]$ $[29\ -11\ -14]$, $T = 4.29$, $k=31$). Some areas correlating with reward prediction errors in both conditions were closely adjacent but non-overlapping at our statistical threshold (**Figure 5.4a**). Most of these activities were located in the lateral portion of the basolateral complex, in an area consistent with the lateral nucleus of the amygdala. We also replicated previous findings of reward prediction error activity in the ventral striatum (which though not the focus of the study was also contained within the field of view of our imaging protocol) during both the reward ($[x\ y\ z]$ $[8\ 7\ -5]$, $T = 4.08$) and avoidance conditions ($[x\ y\ z]$ $[18\ 13\ -11]$, $T = 3.76$) (Gottfried et al., 2003; McClure et al., 2003). Finally, we also tested for an unsigned prediction error in a separate analysis, but did not find evidence for significant correlations with this signal in the amygdala at the time of outcome.

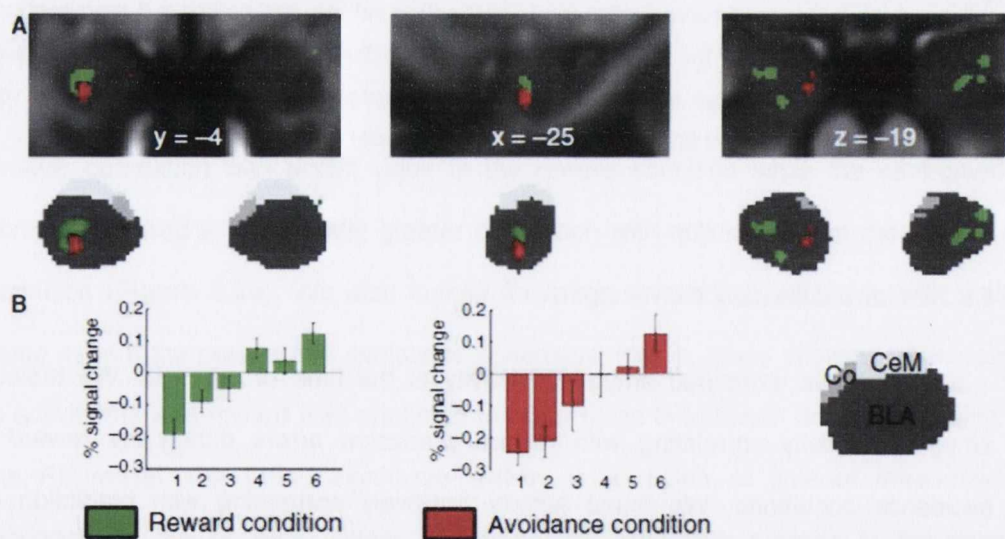


Figure 5.4. Reward prediction error signals. a, BOLD signals correlating with the magnitude of the reward prediction error at the time of outcome were found in the basolateral complex in

both the Reward (in green) and Avoidance (in red) condition. **b**, Plots showing the percent signal change for 6 categories of prediction error values (category 1 representing the lowest prediction error values, and category 6 representing the highest prediction error values) in the Reward (in green) and Avoidance (in red) conditions in the clusters activated.

5.3.2.3 State-value signals

We then examined activity at the time of the initial cue presentation at the beginning of each trial. We looked for regions correlating with the average reward expected at the time of presentation of this initial cue. For this, we assumed that the value of each cue was learned through trial and error in the same way that action values are learned at the time of choice by means of a reward prediction error. We found activity correlating positively with the cue reward value during the avoidance condition in the centromedial complex ($[x\ y\ z]$ $[-18\ -12\ -10]$, $T = 4.29$, $k=30$), such that activity increased the more reward (or less punishment) expected on that trial. 1) No area showed a significant negative correlation with expected reward in the avoidance condition and 2) no region of the amygdala was found to correlate either positively or negatively with expected reward at the time of cue presentation in the reward condition. The fact that we found evidence of this signal in the avoidance but not in the reward condition is difficult to interpret at this stage, and further investigation is needed to elucidate the mechanisms by which estimation uncertainty in reward and avoidance learning operate in the amygdala.

5.3.2.4 Estimation uncertainty signals

Finally, we looked for evidence of a signal widely reported in prior fMRI studies of conditioning in which amygdala activation shows an initial signal increase during the

early stages of learning that subsequently decreases or habituates over the course of learning (Buchel et al., 1998; Davis et al., 2010; Delgado et al., 2008; LaBar et al., 1998). One possible account is that this signal reflects the “estimation uncertainty” of the cue, that is the extent to which attention should be paid to the cue during learning, which can be used to set the rate at which learning occurs. In a Bayesian framework such as that instantiated in a Kalman filter (Dayan et al., 2000), the learning rate of a cue is set using a computation of the variance in the estimation of expected reward. Here, we used a reinforcement learning-based approximation of a Kalman filter to compute the trial by trial uncertainty in the estimate of expected reward for the cues during both the reward and avoidance conditions and generated a model-based prediction for the estimation uncertainty over the course of learning. This signal approximated an exponential decay over the course of learning such that it was high during initial learning on each session when new cues were presented, but then decreased as the estimated predicted reward became more accurate, thus minimizing estimation uncertainty (**Figure 5.5a**). We tested for regions within the amygdala correlating with estimation uncertainty separately for the reward and avoidance conditions. Analogous to the results obtained for the expected value at the time of choice, we found distinct regions of amygdala to be correlated with our estimation uncertainty metric in the reward and avoidance conditions. While a number of clusters in the basolateral complex were correlated with estimation uncertainty in the reward condition ($[x\ y\ z]$ [17 -7 -23], $T = 5.55$, $k=59$; $[x\ y\ z]$ [-31 -5 -17], $T = 4.43$, $k=81$; $[x\ y\ z]$ [-18 -2 -24], $T = 3.96$, $k=39$), a region of dorsal amygdala in the vicinity of the centromedial complex was found to be correlated with estimation uncertainty in the avoidance condition ($[x\ y\ z]$ [-21 -10 -11], $T = 5.66$, $k=46$) (**Figure 5.5b**). Moreover, in an independent ROI analysis, we found a significant subregion $\times$ condition interaction ($F(1,19)=11.94$, $p<0.01$) (**Figure 5.5d**), with the basolateral amygdala being significantly

more involved in representing estimation uncertainty during the reward condition, and the centromedial amygdala being more involved in representing estimation uncertainty during the avoidance condition.

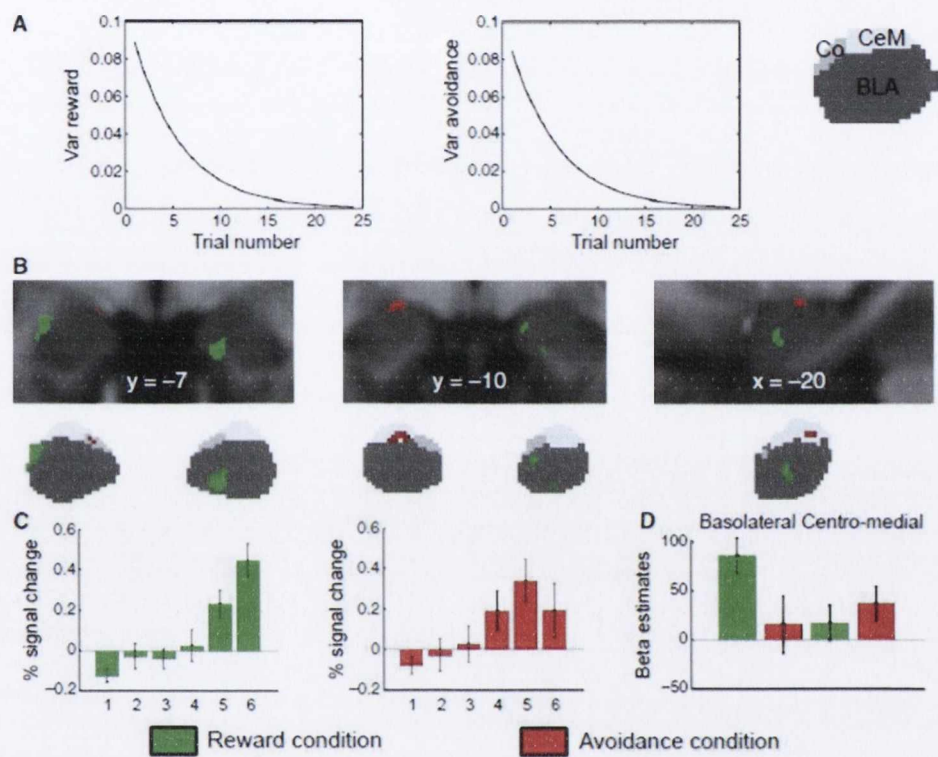


Figure 5.5. Estimation uncertainty signals. *a*, Plots showing the variance (or estimation uncertainty) estimated by our RL based approximation of a Kalman filter averaged across all sessions and subjects in the Reward (left) and Avoidance (right) conditions. Note that these plots show exponential decline over non-contiguous trials, as the trial types were interleaved. *b*, BOLD signals correlating with the magnitude of the estimation uncertainty of the cue were found in the basolateral complex in the Reward condition (in green) and in the centromedial complex in the Avoidance condition (in red). *c*, Plots showing the percent signal change for 6 categories of estimation uncertainty values (category 1 representing the lowest estimation uncertainty values, and category 6 representing the highest estimation uncertainty values) in the Reward (in green) and Avoidance (in red) conditions in the clusters activated. *d*, Beta estimates showing an

interaction between the subregion (basolateral versus centromedial) and the condition (Reward versus Avoidance) ($p < 0.01$).

5.4 DISCUSSION

We provide evidence for at least partly distinct contributions of the basolateral and centromedial amygdala complexes during two fundamental forms of instrumental conditioning: reward and avoidance learning.

Although such a dissociation has never been reported in human studies before to our knowledge, the present results may relate to prior lesion evidence for a differential contribution of basolateral and centromedial complexes in mediating the influence of appetitive and aversive Pavlovian cues on the conditioned potentiation and suppression of instrumental feeding behaviour (Holland et al., 2002; Petrovich et al., 2009). It is possible that there are some similarities in the types of processes being recruited here and in those previous studies. In the present case, visual cues presented at the beginning of each trial served as discriminative stimuli signalling whether or not the subject was in a rewarding or avoidance context, which may be similar to the means by which the Pavlovian stimuli in the conditioned potentiation and suppression paradigms recruited appetitive or aversive motivational processes. Furthermore, the process of selection of an instrumental response in the reward and avoidance conditions might involve the need to promote responding to a previously rewarded action in the reward condition, while suppressing responding to a previously punished action in the avoidance condition.

Expected reward signals were present in the far ventral amygdala in the vicinity of the basolateral nucleus during reward learning, but while we also found expected

value signals during avoidance learning, these signals were located very dorsally within the boundaries of the likely location of the centromedial complex. Although expected reward signals have been reported previously in amygdala in fMRI studies (Elliott et al., 2004; Yacubian et al., 2006), they have hitherto not been localized to a specific nucleus.

In both subregions, the signals we found are “reward expectations”, in that they increased the more reward is predicted to occur and the more a loss predicted to be avoided. Notably, we did not find evidence for aversive-going signals in the avoidance condition at all. Such a result would appear puzzling in the light of considerable evidence of a role for amygdala in aversive learning, especially fear conditioning (Buchel et al., 1998; LaBar et al., 1998; LeDoux, 2003; Schiller et al., 2008). However, monetary loss is a secondary reinforcer and thereby may be processed very differently from more biological reinforcers such as aversive tastes or shock. Indeed, other conventional resolution fMRI studies have also found a lack of increasing activation in amygdala during increasing predictions of monetary loss (Delgado et al., 2008), suggesting that the human amygdala may show signal increases during predictions of some types of aversive reinforcers like shock but not necessarily during anticipation of monetary loss. An important next step will be to assess amygdala responses in these different subregions during learning with other types of aversive reinforcers such as unpleasant tastes or foot shocks, in both Pavlovian and instrumental contexts, to provide learning situations more analogous to that studied in the rodent and non-human primate literature. An alternative explanation for the absence of signals positively correlated with aversiveness could relate to limits in the spatial resolution offered by fMRI (even with a high-resolution imaging protocol). In the event that populations of neurons with rewarding and aversive signalling properties are spatially intermingled

within parts of the amygdala as shown in several electrophysiological studies in rodents and monkeys (Paton et al., 2006; Schoenbaum et al., 1998; Shabel and Janak, 2009), a comparison between appetitive and aversive predicting events might not yield a significant signal change in those areas, even though such neuronal populations are present. In future studies, multivariate fMRI analysis techniques that allow one to detect distributed voxel encoding, might aid in the identification of such populations in human amygdalar subregions, if present.

Interestingly, the signed prediction error signals we found in the basolateral complex for both the reward and avoidance conditions were particularly (though not exclusively) located in the lateral portion of this complex, corresponding to the likely location of the lateral amygdalar nucleus. This area is often considered to be the “gateway to the amygdala” in that it receives inputs from a variety of sensory modalities (LeDoux, 2003, 2007). Plasticity has also been reported to occur within this region of the amygdala during conditioning protocols (Blair et al., 2001; LeDoux, 2007). The presence of reward prediction errors in this region as measured by BOLD could be occurring due to inputs into the amygdala from dopaminergic neurons arising from the ventral tegmental area (Schultz, 2002), which could therefore contribute to modifying plasticity during affective learning. Prediction error signals have previously been reported in basolateral nucleus in non-human primates, as well as in rodents (Belova et al., 2007) although in some cases, the signals were unsigned rather than signed as we observed here (Roesch et al., 2010).

We also observed signals resembling “an estimation uncertainty” computation from a learning process that could be used to set the level of attention allocated to a particular cue, and/or modify the learning rate to that cue. Other possible explanations could be proposed to account for such a decreasing response profile within the

amygdala during learning, such as a more generalized cue-novelty response or arousal response. However, the finding that these signals were located in different regions of the amygdala depending on whether the learning was occurring in a reward or avoidance context is difficult to interpret in terms of a non-specific novelty or arousal signal. The presence of cue-estimation uncertainty signals in the human amygdala is broadly consistent with previous reports that parts of the amygdala mediate the allocation of attention to cues during learning such as in orienting (Gallagher et al., 1990; Roesch et al., 2010).

The main finding of the study is that within the human amygdala, basolateral and centromedial complexes appear to be differentially involved in reward and avoidance learning respectively. The specific involvement of these complexes in these different types of learning might relate to the different patterns of emotional response engendered during the two learning types – while the reward learning condition might engender appetitive preparatory and/or consumatory responses, the avoidance learning condition may evoke greater aversive and/or stress-related responses. Drawing analogies with the animal literature, it is interesting to note that the basolateral amygdala, ventromedial prefrontal cortex and hippocampus are the main sources of input to the ventral striatum, with the densest projections within the nucleus accumbens (Fudge et al., 2002; Haber and Knutson, 2010; Russchen et al., 1985). Because of these convergent fibres, the ventral striatum is considered a “limbic-motor interface”, that is a key structure processing emotional and motivational information further driving action output (Everitt et al., 1999; Haber and Knutson, 2010). Moreover, the basolateral amygdala has been shown to be acting on dopamine-dependent mechanisms of the ventral striatum to mediate the effects of stimulus-reward associations on behaviour (Cador et al., 1989), suggesting that the basolateral amygdala may be involved in

reward learning through its interaction with the ventral striatum. On the other hand, the central nucleus of the amygdala has repeatedly been shown to be uniquely capable of influencing a widespread system of structures including the hypothalamus, midbrain (especially the periaqueductal gray matter) and lower brain stem, via extensive projections shown to be remarkably consistent in rats, cats and monkeys (Price, 2003). These different structures, particularly the periaqueductal gray matter, control and modulate visceral functions through their influence on autonomic function. Therefore, it has been suggested that the central nucleus contributes to the organization of the defense response and the production of fight or flight reactions by integrating the autonomic components of the behaviour (Price and Amaral, 1981). Interestingly, a recent neuroimaging study has shown that distal threat elicited activity in lateral amygdala (consistent with the anatomical location of the basolateral complex) and ventromedial prefrontal cortex (a brain region repeatedly involved in reward processing) while proximal threat elicited activity in dorsal amygdala (consistent with the anatomical location of the centromedial complex) and periaqueductal gray matter (Mobbs et al., 2007). Thus, the differential presence of value signals in these different areas of the amygdala may pertain to a different role for these regions in mediating appetitive and aversive responding respectively. However, it should be noted that the signals found in the avoidance condition in the centromedial complex are not aversive per se, but rather reward-related (increasing the more avoidance is predicted to be successful).

These findings suggest that amygdala subregions might be differentially recruited as a function of different types of learning, namely reward and avoidance learning. It is entirely feasible and indeed likely that reward vs avoidance is not the only dimension on which these amygdala complexes might be differentially engaged. While our findings emphasize partly distinct contributions of the basolateral and centromedial

complexes as a function of the type of instrumental conditioning involved, our computational fMRI analysis revealed that the form of the computations within these two areas appears to be similar. We observed expected reward signals in both of these structures during reward and avoidance learning and we also observed expected uncertainty signals in both of these areas. These findings suggest that these areas are not necessarily engaged in qualitatively different computations, but rather highlight the fact that similar computations in these regions operate on reward and avoidance learning.

The present findings also have important implications for better understanding the neuronal and psychological mechanisms underlying avoidance learning and its relationship to reward learning. Recently, Kim et al. (Kim et al., 2006) reported that medial orbitofrontal cortex was recruited during successful avoidance of a monetary loss in a similar manner to that found during receipt of monetary gain, thereby concluding that avoidance learning depends in part on overlapping neuronal processes to that involved in learning about reward. The present results add to these findings by suggesting that within the amygdala, these two different forms of learning may not rely on entirely overlapping circuitry even if this is the case in the orbitofrontal cortex. However, the fact that the representations were predominantly pertaining to predictions about reward and successful avoidance in both the basolateral and centromedial complexes is broadly consistent with the suggestion that mechanistically, reward and avoidance learning may recruit similar computational processes albeit implemented in partly distinct neural circuits.

5.5 CONCLUSION

Collectively, our findings suggest the existence of a number of distinct computational signals within the basolateral and centromedial complexes, hereby providing new insights into the functional contributions of these amygdalar subregions in associative learning. The adoption of an hr-fMRI approach has allowed us to localize these signals to different amygdalar regions in a manner not hitherto possible in human neuroimaging studies of associative learning. More generally, we show how the hr-fMRI approach can, when combined with quantitative computational modelling, contribute to the circuit level analysis of amygdala function with a level of fidelity hitherto only achievable in animal lesion and neurophysiology studies.

CHAPTER 6 EVIDENCE FOR MODEL-BASED COMPUTATIONS IN THE HUMAN AMYGDALA DURING PAVLOVIAN CONDITIONING

Contemporary computational accounts of instrumental-conditioning have emphasized a role for a model-based system in which values are computed with reference to a rich model of the structure of the world, and a model-free system in which values are updated without encoding such structure. Much less studied is the possibility of a similar distinction operating at the level of Pavlovian conditioning. In the present study, we scanned human participants with fMRI while they participated in a Pavlovian conditioning task with a simple structure. Fitting a model-based algorithm and a variety of model free algorithms, we found evidence at both behavioural and neural levels to support a role for a model-based as opposed to a model-free learning process in the amygdala. These findings support an important role for model-based algorithms in describing the processes underpinning Pavlovian conditioning, as well as providing evidence of a role for the human amygdala in model-based inference.

6.1 INTRODUCTION

Neural computations mediating instrumental conditioning in the mammalian brain have recently been suggested to depend on two distinct mechanisms: a “model-based” reinforcement learning system, in which the value of actions are computed on the basis of a rich knowledge of the states of the world and the nature of the transitions between states, and a “model-free” reinforcement learning system in which the value of actions

are updated incrementally via a reward prediction error without using a rich representation of the structure of the decision problem (Corrado and Doya, 2007; Dayan and Daw, 2008; Doya et al., 2002). Accumulating evidence supports the existence of model-based representations during instrumental conditioning in a number of brain regions, including the ventromedial prefrontal cortex, striatum and parietal cortex (Glascher et al., 2010; Hampton et al., 2006; Li et al., 2011). However, instrumental conditioning is not the only associative learning mechanism in which model-based computations might play a role.

Pavlovian conditioning can also be framed as a model-based learning process, in which the animal begins with a model of the possible structure of the world: the stimuli within it, and sets of possible contingencies that could exist between conditioned stimuli and unconditioned stimuli, as well as assumptions about how these contingencies might change over time. In essence, these sets of possible contingencies constitute latent causes, and learning within the system, can be thought of in terms of determining statistical evidence for which latent cause out of the set of possible latent causes describes the current causal structure of the environment (in terms of which stimulus causes which unconditioned stimulus), as well as for determining whether or when the relevant causal processes have changed as a function of time. Model-based approaches to classical conditioning to date have tended to use Bayesian methods to yield inference over structure (Courville et al., 2006). Consequently, such models in addition to encoding probabilistic beliefs about possible causes, also incorporate estimates of the degree of precision in the model predictions as to which causal structure is currently in place, and about when the causal structure has changed given the observable evidence.

Very little is known however, about the extent to which such model-based algorithms are actually implemented in the brain during Pavlovian conditioning. The aim of the present study was to address this question using a computational fMRI approach. We scanned human participants with fMRI while they participated in a Pavlovian conditioning procedure that incorporated a rich structure. The task involved both an appetitive and an aversive conditioning procedure during which participants were presented with cues leading to the probabilistic delivery of affectively pleasant or affectively neutral liquid flavor outcomes or else to affectively unpleasant or affectively neutral outcomes respectively. Each session began with two novel cues associated with either the affectively significant flavor liquid or the neutral liquid. These associations reversed after an unpredictable number of trials. After another unpredictable number of trials following reversal onset, a novel pair of cues was introduced, whose associations reversed following the same rule. Likewise, a third novel pair of cues was introduced wherein associations with the flavored liquids also reversed once. The hidden structure in this task was knowledge that for each novel pair of conditioned stimuli, the contingencies within a given conditioning session would reverse.

To construct a model-based learning algorithm, we devised a Hidden Markov model (HMM) incorporating knowledge that one cue would be causally associated with the affectively significant outcome, while the other cue would be causally associated with the affectively neutral outcome, as well as knowledge that the contingencies would reverse. This model then performed Bayesian inference about the likely causes of experienced outcomes using the cue stimuli and outcomes as input. We compared this model against two influential “model-free” algorithms for Pavlovian conditioning: a Rescorla-Wagner (RW) learning rule in which signed prediction errors based on the discrepancy between expected and actual outcomes are used to update associative

strengths for Pavlovian cues (Rescorla, 1972), as well as a Pearce-Hall (PH) learning rule in which the absolute degree of surprise (unsigned prediction error) engendered by an unpredicted delivery of an unconditioned stimulus is used to modify the associability assigned to a particular conditioned stimulus (Pearce and Hall, 1980). The conditioning task that the participants experienced is by virtue of its design, ideally placed to enable the predictions of these two competing modeling approaches to be compared and contrasted. The model-based algorithm should exploit the structure and update expectations based on when contingency reversals are anticipated to occur, whereas the model-free algorithms would not expect reversals to occur and would therefore update. In order to test for model-based signals in the brain during Pavlovian conditioning, we focused in particular on the amygdala, a structure that has been heavily implicated in Pavlovian conditioning on the basis of a large body of both animal and human studies (Buchel and Dolan, 2000; Delgado et al., 2006; Fanselow and LeDoux, 1999; Johansen et al., 2011). In order to enable us to obtain signals from this region with the best possible fidelity, we used a high-resolution fMRI protocol in which we acquired images with more than 4 times the resolution of a standard 3mm isotropic scan, alongside an amygdala specific normalization procedure (Prevost et al., 2011). We then generated regressors corresponding to different components of our computational models (including expected value, precision, prediction error and associability) and correlated these against fMRI data acquired from the amygdala separately for the appetitive and aversive conditioning procedures. We then performed model comparison between the model fits to both the behavioural and neuroimaging data. We hypothesized that the model-based algorithm would account better for both behavioural and fMRI data acquired during both the appetitive and aversive conditioning phases than would the models of Pavlovian conditioning which do not contain such structured knowledge.

6.2 EXPERIMENTAL PROCEDURES

6.2.1 Subjects

Nineteen right-handed subjects (8 females) with a mean age of 22.21 ± 3.47 participated in the study. All subjects were free of neurological or psychiatric disorders and had normal or correct-to-normal vision. Written informed consent was obtained from all subjects, and the study was approved by the Trinity College School of Psychology Research Ethics committee.

6.2.2 Task description

Subjects participated in a Pavlovian task where they had to learn associations between different cues (fractal images) and a pleasant (blackcurrant juice [Ribena, Glaxo-Smithkline, UK]), affectively neutral (artificial saliva made of 25mM KCl and 2.5 mM NaHCO₃) or unpleasant (salty tea made of 2 black tea bags and 29g of salt per liter) flavor liquid. The task consisted of two sessions lasting approximately 22 minutes each. Each session was composed of 120 trials, leading to a total of 240 trials. In one of the sessions, subjects underwent an appetitive Pavlovian conditioning procedure whereby they were presented with cues leading to the subsequent delivery of either the pleasant flavor, or the affectively neutral one, while in the other aversive conditioning session, subjects underwent an aversive conditioning procedure whereby they were presented with cues leading to the subsequent delivery of either the unpleasant flavor stimulus, or else the affectively neutral stimulus. The rationale for including the appetitive and aversive conditioning procedures in separate sessions as opposed to including both conditions intermixed within the same sessions was to avoid contrast effects observed in prior behavioural piloting whereby cues signaling the aversive

outcome tended to overwhelm cues signaling the pleasant one such that both the pleasant and the neutral cue stimuli were viewed as relief stimuli (contrasted against the aversive outcome) (Seymour et al., 2005). Performing the appetitive and aversive conditioning procedures in separate sessions ensured robust behavioural conditioning in both the appetitive and aversive cases and largely avoided contrast effects between the appetitive and aversive conditions.

For both sessions, on each trial, a cue was displayed randomly on either the left or right side of a fixation cross for 4 seconds. Following a well-established Pavlovian conditioning protocol (Gottfried et al., 2002, 2003; O'Doherty et al., 2004), subjects were also instructed to indicate on which side of the screen the cue was presented by means of pressing the laterally corresponding button on a response box, yet they were also instructed that the subsequent outcomes were not contingent on their responses. This serves two purposes: it allows one to monitor the extent to which participants are paying attention to the cues on each trial, as well as offering a response time measure which can serve as an index of conditioning. The offset of the cue (after 4 secs) was followed by delivery of one of the liquid flavor stimuli with a probability of 0.6, or else no liquid stimulus was delivered. The next trial was triggered following a variable 2-11 secs inter-trial interval.

At the beginning of each session, subjects were presented with two novel fractal cues (not seen before in the course of the experiment): which we will denote as cue 1 and cue 2. In the Appetitive session, cue 1 predicted the subsequent presentation of the pleasant liquid 60% of the time (or no liquid delivery 40% of the time), while in the Aversive session cue 1 predicted the delivery of the aversive liquid 60% of the time (or no liquid delivery 40% of the time). Cue 1 and Cue 2 trials were presented in a randomly intermixed order. After 16 trials (8 trials of each type), a reversal of the cue-

outcome associations was set to occur with a probability of 0.25 on each subsequent trial. The probabilistic triggering of the reversal after the 16th trial ensured that the onset of the reversal was not fully predictable by subjects. Once a reversal was triggered, cue 1 no longer predicted the appetitive or aversive outcome but instead was associated with delivery of the neutral outcome, while cue 2 now predicted the appetitive or aversive outcome. After another 16 trials (8 trials of each type) following the onset of the reversal, another event was triggered to occur with probability 0.25 on one of the subsequent trials: this time instead of a reversal, a completely novel pair of stimuli was introduced. One of these, cue 3, was now paired with the appetitive or aversive outcome, while cue 4 was now paired with the neutral outcome. These new cues were presented for a further 16 trials, and followed again after a probabilistic trigger of $p=0.25$ on each subsequent trial with a reversal of the associations. After the reversal, a new set of cues were introduced according to the same probabilistic rule and this was followed again by a reversal. Thus in total, 3 unique pairs of stimuli were used in each session and each of these pairs underwent a single reversal (Figure 6.1). A completely different set of cues were used for each session, so that subjects experienced a total of 6 pairs of fractal stimuli throughout the whole experiment.

Within each session, the presentation order of the affective and neutral cue presentations was randomized throughout, with the one constraint that the cue predicting the neutral tasting liquid delivery had to be delivered twice every four trials. This ensured that the appetitive and neutral cues, and aversive and neutral cues were approximately evenly distributed in their presentation throughout the appetitive and aversive sessions respectively. All fractal images were matched for luminance. The order of the sessions was counterbalanced across subjects so that half of the subjects

started the experiment with the appetitive session and half of the subjects with the aversive session.

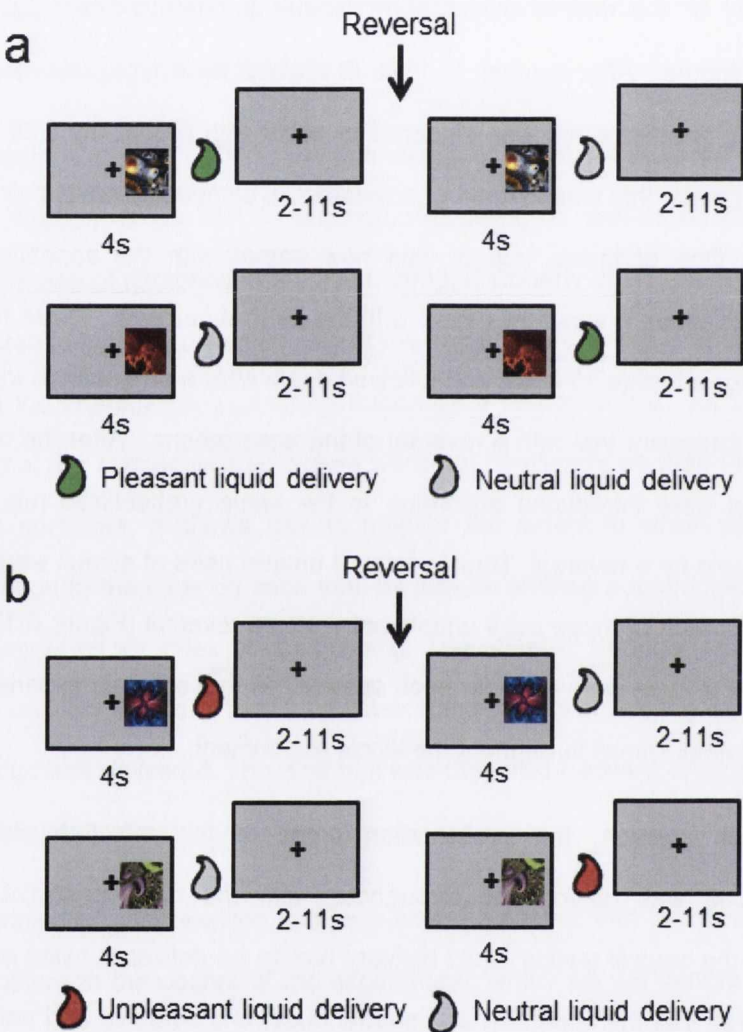


Figure 6.1. Appetitive versus aversive Pavlovian learning task. Sequence and timing of events in the appetitive (a) and aversive (b) sessions. On each trial, a cue was presented on one side of the screen for 4s, followed by some liquid delivery 60% of the time. The trial ended with a 2-11s inter-trial interval. Each session started with the

presentation of cue 1 and cue 2, leading 60% of the time to a pleasant or a neutral liquid delivery in the appetitive session or an unpleasant or a neutral liquid delivery in the aversive session. After a number of trials, a reversal occurred so that cue 1 now led to the liquid associated with cue 2, and cue 2 led to the liquid associated with cue 1. Subsequently, a new pair of cues was presented, which also reversed after a number of trials. In total, three new pair of cues were presented, and each of these pairs reversed once.

6.2.3 Subject instructions

Before the conditioning session, subjects received the following task instructions:

“In each trial, an image will appear on the screen and may be followed by some liquid delivery. There are six different images per session. Each image will lead to either a pleasant, neutral or unpleasant tasting liquid. You will have to learn these associations. However, during the experiment, this may change (or reverse), making image 1 associated with the liquid of image 2 and image 2 associated with the liquid of image 1. This reversal may actually happen more than once during the experiment and you have to fully pay attention and realize that it has happened. These cues may change during the experiment, so that you will have to learn these associations again with these new cues (which may also reverse).

At the beginning of each trial, the image will either appear on the left or right side of the screen. You will have to press the left button of the response pad if the image appears on the left side, or the right button if it appears on the right side. It is important that you press the button because we need to record your response times, although the trial will carry on if you don't press any button.

At the beginning and end of each session, we will ask you to rate different images and liquids. You will also have to rate these images in the middle of each session.”

6.2.4 Apparatus

The pleasant, neutral and unpleasant tasting liquids were delivered by means of three separate electronic syringe pumps positioned in the scanner control room. These pumps pushed 1 mL of liquid to the subject’s mouth via ~10 m long polyethylene plastic tubes, the other end of which were held between the subject’s lips like a straw, while they lay supine in the scanner.

6.2.5 Behavioural measures

6.2.5.1 Affective evaluations of the fractal images and liquids

Participants were asked to provide subjective ratings indicating their perceived subjective hedonic evaluation for each of the 6 pairs of fractal images that were displayed. This was done during the experiment before each session, in the middle of each session (during the scanning) and at the end of each session, by presenting a picture of the fractal alongside an instruction to rate the fractal for its pleasantness on a scale going from 1 (do not like at all) to 4 (strongly like). These ratings could therefore provide a behavioural measure of evaluative conditioning (Bray et al., 2008) at three different time points throughout the experiment. Furthermore, before and after the appetitive session, the pleasant and neutral liquids were rated for their subjective pleasantness using a scale ranging from -5 (very unpleasant) to +5 (very pleasant), and

similarly the aversive and neutral liquids were rated before and after the aversive session.

6.2.5.2 Preference ranking test

Before the experiment started and after the experiment was over, participants were asked to make binary choices indicating their relative preferences for each of 16 different fractals (12 of which were included in the experiment; 6 each in the appetitive and aversive sessions respectively; while 4 of the fractals were not featured in either session). Each of the 16 fractals was paired with each other fractal. This test allowed us to estimate a preference ranking for each of the fractals, thereby potentially providing an additional even more direct behavioural metric of evaluative conditioning beyond the pleasantness ratings.

6.2.5.3 Pupillary dilation

Pupil diameter was continuously measured during scanning using an MRI compatible integrated goggle and infrared eye tracking system (NordicNeuroLab AS, Bergen, Norway). Pupil reflex amplitude has been shown to be modulated by arousal level and can therefore be used as a physiological index of conditioning (Bitsios et al., 2004; Bray et al., 2008; Seymour et al., 2007). Pupil measurementss could not be taken from 9 participants because space constraints within the head-coil alongside variations in head size meant that in some individuals the eye-tracker could not fit them comfortably.

6.2.5.4 Fluctuations in respiration and heart rate

Estimates of heart rate and respiration were recorded using a pulse oximeter positioned on the forefinger of subjects' left hand and a pressure sensor placed on the umbilical region. The time courses derived from these measures were used as a further physiological index of conditioning as well as being used separately to remove physiological noise from the fMRI data analysis (see fMRI data analysis).

6.2.6 Data Acquisition

Functional imaging was performed on a 3T Philips scanner equipped with an 8-channel SENSE (sensitivity encoding) head coil. Since the focus of our study was on the amygdala, we only acquired partial T2*-weighted images centered to include the amygdala while subjects were performing the task. These images also encompassed the ventral part of the prefrontal cortex, the ventral striatum, the insula, the hippocampus, the ventral part of the occipital lobe and the upper part of the cerebellum (amongst other regions). Nineteen contiguous sequential ascending slices of echo-planar T2*-weighted images were acquired in each volume, with a slice thickness of 2.2 mm and a 0.3 mm gap between slices (in-plane resolution: 1.58 x 1.63 mm; repetition time (TR): 2000 ms; echo time (TE): 30 ms; field of view: 196 x 196 x 47.2 mm; matrix: 128 x 128). A whole-brain high-resolution T1-weighted structural scan (voxel size: 0.9 x 0.9 x 0.9 mm) and three whole-brain T2*-weighted images were also acquired for each subject. To address the problem of spatial EPI distortions which are particularly prominent in the medial temporal lobe (MTL) and especially in the amygdala, we also acquired gradient field maps. To provide a measure of swallowing motion, a motion-sensitive inductive coil was attached to the subjects' throat using a Velcro strap. The

time course derived from this measure was used as a regressor of no interest in the fMRI data analysis. Finally, to account for the effects of physiological noise in the fMRI data, subjects' cardiac and respiratory signals were recorded with a pulse oximeter and a pressure sensor placed on the umbilical region and further removed from time-series images. We discarded the first 3 volumes before data processing and statistical analysis to compensate for the T1 saturation effects.

6.2.7 Preprocessing

See section 3.2.7.

6.2.8 Computational model analysis

To test whether amygdala activity was better explained by model-based or model-free learning algorithms, we correlated brain activity in this region with expected value signals estimated by a number of different computational models. In 1) model-free learning algorithms, the agent is surprised when a reversal occurs and starts learning again after it happens, whereas in 2) model-based learning algorithms, the agent expects the reversal and considers it as resolution of uncertainty and does not need to relearn. The two modes of learning are diametrically opposed in the current task, therefore allowing us to test whether amygdala is tracking model-based or model-free computations.

6.2.8.1 Model-based learning algorithms

- HMM with dynamic expectation of change

For the model-based learning algorithm, we used a Hidden Markov Model (HMM). In this HMM, the inferred state of the environment is defined in terms of an association between cues and outcomes. There are three possible liquid outcomes (pleasant, neutral, and unpleasant) and two cues on any given trial. The state values S_t are the possible combinations of cues and outcomes, for example $S_t = (\text{cue 2}, \text{neutral liquid})$. Although the subjects were unaware that pleasant and unpleasant outcomes could not be delivered concurrently, this possible state value was omitted since it did not affect the results of the analyses. We also incorporated a binary-valued variable H in this HMM. The values of this hidden node determine whether ($H = 1$) or not ($H = 0$) the subject is expecting a reversal. A third random variable O represents the observed cue-outcome combination (see Figure 6.2b for a simple graphical representation of the model).

The transition probabilities of the reversal variable H are:

$$P(H_t|H_{t-1}) = \begin{pmatrix} 1 - \alpha & \alpha \\ 0 & 1 \end{pmatrix}$$

Variable values are enumerated along the row and column axes. Each entry of the matrix represents the probability of moving from one value on trial $t - 1$ (rows) to another on trial t (columns). At position (1,2), the α parameter is the probability of moving to the state of expecting a reversal ($H = 1$) from the $H = 0$ state. Once a subject begins expecting a reversal, they do not switch back. This is encoded in the asymmetry of the transition matrix. The time evolution of H represents a subject's growing expectation of a reversal in the cue-outcome association. After the presentation of a

novel pair of cues, H is set to the zero state. The transitions for the state variable S are conditionally dependent on the reversal variable:

$$P(S_t|S_{t-1}, H_t) = \begin{pmatrix} 1 - \beta & \beta \\ \beta & 1 - \beta \end{pmatrix}$$

State reversals are inferred with a non-zero probability β when H is in the reversal expectation state ($H_t = 1$), otherwise $\beta = 0$ and $P(S_t|S_{t-1}, H_t = 0)$ is the identity matrix. The posterior probability distribution $P(S_t)$ over the state values on trial t is determined by the prior state probability distribution $P(S_{t-1})$, the cue-outcome observation O_t , and the state transition probabilities:

$$\mathbf{Prior}(S_t) = \sum_{S_{t-1} \text{ states}} \sum_{H_t \text{ states}} P(S_t|S_{t-1}, H_t) P(H_t) \mathbf{Posterior}(S_{t-1})$$

$$\mathbf{Posterior}(S_t) = \frac{P(O_t|S_t) \mathbf{Prior}(S_t)}{\sum_{S_t \text{ states}} P(O_t|S_t) \mathbf{Prior}(S_t)}$$

The prior over the state values at the beginning of a new set of cues is uniform. Beliefs are updated based on the likelihood of observing an outcome for a given cue and assuming a state such as “cue j is rewarding and this is likely to reverse soon”. For instance, if no reward is observed for cue j , then this state is given less credence because the likelihood that this occurs is low (0.4), and the mathematical expectation of the reward for cue j is decreased. Significantly, expectations for the other cue are updated simultaneously, even if it is not implicated in the current trial. This is because a lower chance for the state “cue j is rewarding and this is likely to reverse soon” implies that the state “the other cue is rewarding and this is unlikely to reverse soon,” is more likely, and hence, the mathematical expectation of the reward upon presentation of the other cue increases.

The expected reward Q_j when presented with a given cue j is

$$Q_j(t) = E[R|cue\ j, trial\ t] = \sum_{R\ rewards} \sum_{S_t\ states} R P(R|S_t, cue\ c) P(S_t)$$

The reward R takes the values -1, 0, 1 for unpleasant, neutral, and pleasant rewards respectively. Here, “E” denotes the mathematical expectation operator. Because the forecast is based on the mathematical expectation, it is unbiased. This means that the forecast is correct on average for all possible outcomes given a specific history of past rewards for both cues.

Confidence in, or precision about, the identity of the current state can be measured by the extent to which there are differences in the posterior probabilities of the possible states given past experience and the cues presented. When these differences are high, one posterior probability is necessarily high, and hence, precision is high. Conversely, if all posterior probabilities are the same, precision is lowest.

The agent anticipates surprise more and more as reversals take longer to occur, and hence, precision in the true state decreases. We model the increase of the chance that a reversal happens over time using the H node in the HMM. The H node needs to switch from $H = 0$ to $H = 1$ before a reversal is inferred and the likelihood of this switch increases the longer the HMM has evolved. As more and more trials with no reward are experienced, the H node inputs a growing uncertainty about the identity of the current state into the HMM (since a reversal may have occurred in the absence of a rewarding outcome). Every time a new pair of cues is presented, precision is low but increases dramatically when the agent knows what particular state they are in (i.e. what the cue-liquid association is). Precision lowers again until the agent knows that a reversal has occurred, after which precision increases again.

6.2.8.2 Model-free learning algorithms

- Rescorla Wagner model.

In the Rescorla Wagner (RW) model, the new expected value at trial $t + 1$ for a given cue is based on the sum of the current expected value and the prediction error between the reward obtained and the expected value at time t , weighted by the learning rate (Rescorla, 1972):

$$Q_j(t + 1) = Q_j(t) + \alpha \cdot (R(t) - Q_j(t))$$

When j is a given cue, α is the learning rate with a range $0 \leq \alpha \leq 1$, and $R(t)$ is the reward received on the current trial. If the valenced (pleasant or unpleasant) liquid was obtained on the current trial, $R(t) = 1$, else $R(t) = 0$. Hence there is one free parameter in this model, α ,

-Pearce Hall model.

This model differs from the Rescorla Wagner model (RW) in that it introduces an associability component and allows the effectiveness of the reinforcer to remain constant throughout conditioning. The associability values estimated by this model will decrease as the consequences of the conditioned stimulus become accurately predicted (Pearce and Hall, 1980). The expected values $Q(t)$ of a given cue were updated according to:

$$Q_j(t + 1) = Q_j(t) + S \cdot |R(t - 1) - Q_j(t - 1)| \cdot R(t)$$

When j is a given cue, S is a free parameter governing the intensity of the CS, and $R(t)$ is the reward received on the current trial. If the valenced (pleasant or unpleasant) liquid was obtained on the current trial, $R(t) = 1$, else $R(t) = 0$. In the Pearce Hall model (PH),

the new expected value at trial $t + 1$ for a given cue is based on the sum of the current expected value and the product of the absolute value of the difference between the outcome obtained on the previous trial and the expected reward on the previous trial, and the outcome obtained on the current trial; this product is weighted by the free parameter. Hence there is one free parameter in this model, S .

-HMM model with static expectations of change

In order to further test whether amygdala activity is tracking precision signals from a model-based algorithm as opposed to more generically tracking precision signals computed in a model-free manner, we used a simpler version of the HMM described above where precision signals resemble more closely what a "model-free" algorithm would estimate. In this version of the HMM, H is always set to the $H = 1$ state and thus the chance of a reversal happening is constant over time. As a result, in this HMM, precision starts low every time a new pair of cues is presented and increases substantially when the agent knows in which state they are, but because the chance of a reversal occurring does not increase over time, the precision remains high through the rest of the learning with that cue until a new pair of cues is introduced. In other words, there is no decrease in precision related to the anticipation of a change in the contingencies (which would come from having a model of when the contingencies are predicted to reverse), but instead a decrease in precision occurs only once a contingency change has occurred and been detected through trial and error experience (hence the algorithm is essentially model-free). Although the precision signals generated by our "model-based" and "model-free" HMM are very different, the expected reward signals from both signals are strongly correlated.

-Baseline model

Our baseline model simply assumes that rewards occur completely at random and no learning takes place. Hence, expected values for all trials are kept at a constant value of 0.5.

6.2.8.3 Model comparison on behavioural data

To perform a formal model comparison on the behavioural conditioning data, we used the trial-by-trial reaction time data (measuring the length of time taken on each trial for participants to press a button to indicate which side of the screen the Pavlovian cue stimulus had been presented). Many previous studies have shown that changes in RTs to a Pavlovian cue are correlated with changes in associative encoding between cues and behaviourally significant outcomes (Bray and O'Doherty, 2007; Gottfried et al., 2003; O'Doherty et al., 2004). For each session separately, we log transformed and adjusted the RT data to account for a linear trend in RTs over time independently of trial type, as well as to remove the effects of changes in reaction time related to switching responses from one side of the screen to the other. This was done by regressing the log transformed RTs against a matrix containing a column of ones, a column accounting for the linear trend over time and a column indicating whether participants switched their response from left to right or vice versa between the current and previous trial using the function `regress` in Matlab.

Using the same function, we then regressed these adjusted response times against the expected values generated by our 'model-based' HMM, our 'model-free' RW and PH models and our baseline model. (For the baseline model, a small amount of noise was added to each expected value in order to compute the regression; without any noise the

regression would not be calculable). This 2nd regression analysis was run for each of these models, and cycled through all the possible learning rate parameters for the RW model, and CS intensity parameters for the Pearce-Hall model between 0 and 1, with increments of 0.001. This method returned Sum Squared Error (SSE) values for each of these parameter values thereby allowing us to obtain the best fitting value for the free parameter for the appetitive and aversive sessions (i.e. the free parameter associated with the lowest SSE value). In order to compare the goodness of fit between these four different models, we converted the best SSE value of each session (appetitive and aversive) and each model into a Bayesian information criterion (BIC) value. The BIC adds a penalty proportional to the number of additional free parameters to the SSE value of each model, depending also on the number of degrees of freedom which in this case, is the total number of trials per session across all subjects (Schwarz, 1978). Using this procedure, we found that in both the appetitive and aversive sessions, the 'model-based' HMM outperformed the baseline model, and the baseline model outperformed both the 'model-free' RW and PH models. Therefore, the 'model-based' HMM best fit our behavioural data, whereas the best fitting RW and PH models did not fit our behavioural data better than a random model. Hence, unlike RW and PH, the 'model-based' HMM predicted RTs better than chance performance. Note that we did not regress the expected values generated by our simple HMM since they were highly correlated with that of our 'model-based' HMM.

a

RW	$Q_j(t+1) = Q_j(t) + \alpha(R(t) - Q_j(t))$
PH	$Q_j(t+1) = Q_j(t) + S \times R(t-1) - Q_j(t-1) \times R(t)$
HMM	$Q_j(t) = E[R cue\ j, trial\ t]$ $= \sum_{R\ rewards} \sum_{S_t\ states} R P(R S_t, cue\ c) P(S_t)$

b

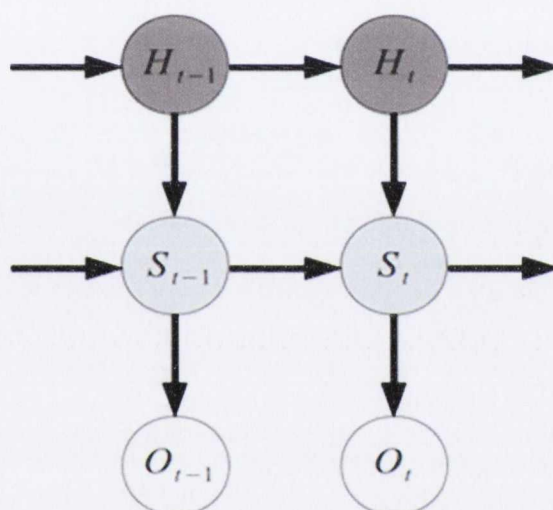


Figure 6.2. a, Computational models used to estimate expected reward on each trial (Qj). The expected rewards generated by the model-free learning algorithms (Rescorla-Wagner (RW) and Pearce-Hall (PH) were compared against a model-based learning algorithm (Hidden Markov Model or HMM) at both the behavioural and neural levels. b, Graphical model representation of the Bayesian HMM.

6.2.9 fMRI data analysis

The event-related fMRI data were analyzed by constructing sets of δ (stick) functions at the time of cue presentation and at the time of outcome for the appetitive and aversive sessions. For our main GLM (illustrated in figures 6.4 and 6.5), additional regressors were constructed by using the expected values and the precision values generated by the model-based HMM as modulating parameters at the time of cue presentation. In order to compare model-based versus model-free learning algorithms in the amygdala, we ran thirteen additional GLMs. To avoid biasing our model comparison using an a priori free parameter, we estimated expected rewards using the full range of plausible free parameters (i.e 0.1 to 0.6 by increment of 0.1) for both RW and PH. For RW, the regressors were similar to our model-based HMM except that we did not have a regressor for precision which is not estimated by RW, and we added a modulating parameter for prediction error at the time of outcome. The regressors used in the six GLMs computed using PH models were the same as the ones used in our model-based HMM, except that the precision modulating parameter was replaced with an associability modulating parameter at the time of cue presentation. Finally, we ran a "model-free" HMM GLM using the same regressors as for our model-based HMM. All of these regressors were convolved with a canonical hemodynamic response function (HRF). The six scan-to-scan motion parameters derived from the affine part of the realignment procedure were included as regressors of no interest to account for residual motion effects. To account for motion of the subjects' throat during swallowing, we added a regressor of no interest for swallowing motion. Finally, we also included thirteen additional regressors to account for physiological fluctuations (4 related to heart rate, 9 related to respiration) which were estimated using the RETROICOR algorithm (Glover et al., 2000). Six of the 38 (2 sessions x 19 subjects) log files could not be used

to estimate these regressors due to a technical problem during data collection, and the missing physiological regressors were simply omitted for those sessions. All of these regressors were entered into a general linear model and fitted to each subject individually using SPM5. The resulting parameter estimates for regressors of interest were then entered into second-level one sample t-tests to generate the random-effects level statistics used to obtain the results shown in figures 6.4 and 6.5. All reported fMRI statistics and p values arise from group random-effects analyses. We present our statistical maps at a threshold of $p < 0.005$, corrected for multiple comparisons at $p < 0.05$. To correct for multiple comparisons, we first used the 3dFWHMx function in AFNI to estimate the intrinsic smoothness of our data, within the area defined by a mask corresponding to our amygdala template. We then used the AlphaSim function in AFNI to estimate via Monte Carlo simulation an extent threshold for statistical significance that was corrected for multiple comparisons at $p < 0.05$ for a height threshold of $p < 0.005$ within the amygdala ROI.

6.2.10 Model comparison on BOLD data

In order to test whether the amygdala acts according to model-based or model-free learning algorithms, we used a Bayesian model selection procedure (BMS) to test which expected value signals estimated by model-based versus model-free learning algorithms better accounted for amygdala activity (Solano-Castiella et al., 2010). For both the appetitive and aversive sessions, we included in this model comparison voxels within a 4mm sphere centered on the peak voxel of the amygdalar activity correlating with expected value signals for the HMM using the leave-one out method, thereby avoiding a non-independence bias in the voxel selection (Kriegeskorte et al., 2009).

Using the `spm_BMS` function in SPM8, we compared expected value signals across all model-based (HMM) and model-free (6 RW and 6 PH using free parameters in the range 0.1-0.6 by increment of 0.1) models separately for the appetitive and aversive sessions.

We used a similar approach to compare neural activity pertaining to precision signals estimated by our “model-based” and “model-free” HMMs. The difference between these two HMMs is that the “model-based” HMM does not allow for a reversal without moving from a “non-reversal state” to a “possible reversal state”. As a consequence, the precision values generated by these models are clearly distinguishable and thus easily comparable using a BMS (whereas the estimated expected rewards are strongly correlated). Again, we included in this model comparison voxels within a 4mm sphere centered on the peak voxel of the amygdalar activity correlating with precision signals for the “model-based” HMM using the leave-one out method. Here, we compared activity correlating with precision signals between the “model-based” and “model-free” HMM separately for the appetitive and aversive sessions (see Results section for the exceedance probabilities).

6.2.11 ROI analyses

Functional regions of interest (ROIs) were defined using the MarsBaR toolbox (<http://marsbar.sourceforge.net/>). Beta estimates were extracted for each subject from the functional clusters of interest as they appeared on the statistical maps of a given contrast using the leave-one out method to avoid a non-independence bias. They were then averaged across subjects to plot expected reward (Figure 6.4b) and precision

(Figure 6.5b) according to 3 categories (category 1 corresponding to the lowest values and category 3 corresponding to the highest values).

6.3 RESULTS

6.3.1 Behavioural results

6.3.1.1 Affective ratings for the liquid outcomes

Subjects were asked to give subjective ratings of the pleasant and neutral tasting liquids before and after the appetitive session and of the unpleasant and neutral tasting liquids before and after the aversive session. The pleasant, neutral and unpleasant tasting liquids (unconditioned stimuli or USs) were reported to be highly pleasant, neutral and unpleasant by subjects as indicated by their ratings averaged across before and after conditioning (figure 6.3a). There was no significant difference in the pleasantness ratings of any of the liquid outcomes before and after conditioning (paired t-tests, all $p > 0.05$).

6.3.1.2 Revealed preference rankings for the cue stimuli

Subjects made binary preferences between the visual cues used in the conditioning protocols before and after the experiment (Figure 6.3b). Subjects showed increased preference rankings for the cues displayed in the appetitive session (averaging across both CS+ and CS- cues as both were paired with reward and neutral outcomes over the course of the experiment due to the reversal) after as compared to before the experiment ($p < 0.001$). Furthermore, the set of cues used in the aversive sessions

showed a significant decrease in their relative preference rankings ($p < 0.001$). Preference rankings for the control cues (cues not included in either the appetitive or aversive conditioning sessions) showed no significant changes from before to after the experiment. These results indicate that while the cues displayed in the appetitive session have acquired an increased positive value, those displayed in the aversive session have acquired a negative value; indicating that subjects showed a modulation in their affective responses to the cue stimuli as a function of the context in which these stimuli had been conditioning (appetitive vs aversive).

6.3.1.3 Pleasantness ratings for the cue stimuli

We also obtained pleasantness ratings from subjects while in the scanner during the conditioning procedure. In the middle of the appetitive session, a few trials after a new pair of cue was presented, subjects rated the cue paired with the pleasant liquid significantly higher than the cue paired with the neutral liquid ($p < 0.01$) (figure 6.3c). Subjective ratings were obtained at the end of the appetitive session, hence following reversal of the last pair of cues and although they still rated the cue paired with the pleasant liquid higher than the one paired with the neutral liquid, this difference was not significant. Similarly, in the aversive session, the cue paired with the unpleasant liquid was rated significantly higher than the cue paired with the neutral liquid a few trials after a novel pair of cue was presented ($p < 0.01$) but not after a reversal had occurred (figure 6.3d).

6.3.1.4 Pupil diameter and blink rate changes as a function of conditioning

We recorded pupil diameter, an automatic measure of arousal previously shown to provide a measure of conditioning (Bitsios et al., 2004; Bray et al., 2008; Seymour et al., 2007). We found a significantly smaller amplitude in pupil diameter for trials where the cue was predictive of the pleasant liquid (appetitive condition) as compared to trials where the cue was predictive of the neutral liquid (neutral condition) ($p < 0.05$) in a time window of 0.8-1.5s after cue onset where amplitude changes in pupil diameter have previously been reported (Seymour et al., 2005) in the 10 subjects from which we obtained pupil amplitude measures (figure 6.3e). A higher degree of arousal (significantly smaller peak amplitude) would have been equally expected when subjects saw cues predictive of the aversive liquid; however, reliable analysis of amplitude in pupil diameter for these trials was prevented by the prolonged blinking elicited by these aversive cues. Given that blinking is also a conditioned response, we looked for evidence of blinking in the aversive condition as opposed to the neutral condition. We found significant differences in blinking between the aversive and neutral conditions during the first second after cue onset and last second before cue offset (paired t-tests, $p < 0.05$) as well as at the time of liquid delivery and swallowing (paired t-tests, $p < 0.01$).

6.3.1.5 Changes in pulse and respiration patterns as a function of conditioning

Pulse (an estimation of heart rate) and respiration were also continuously recorded during scanning. Time courses averaged across trials from the appetitive and aversive conditions are plotted in figure 6.3f,g. Pulse amplitude was maximal two seconds after cue onset and markedly decreased three seconds after cue onset in the aversive condition whereas the reverse pattern was observed in the appetitive condition (figure

6.3f). These amplitudes were significantly different at peaks from both the aversive (2s) and appetitive (3s) conditions ($p < 0.05$). In the aversive condition, subjects learned to inspire before cue offset and expire at the time of the aversive liquid delivery (figure 6.3g). In contrast, subjects expired before cue offset and inspired at the time of the appetitive liquid delivery in the appetitive condition. The amplitudes between the appetitive and aversive conditions were significantly different both before cue offset (3.5s) and at the time of liquid delivery (4.5s) ($p < 0.05$).

6.3.1.6 Model Comparison on Behavioural Data

We used Bayesian information criterion (BIC) to compare the goodness of fit of the HMM against the baseline model and the model-free algorithms on the basis of trial by trial variation in reaction times. We found that the HMM model fit better than each of the other models including the baseline model, indicating that this model was providing the best account of trial by trial variation in conditioning as reflected in reaction times. On the other hand, neither the RW nor the PH learning rules provided a better fit to the data than did the baseline model, suggesting that these algorithms cannot account for changes in reaction time as a function of conditioning any better than a random actor.

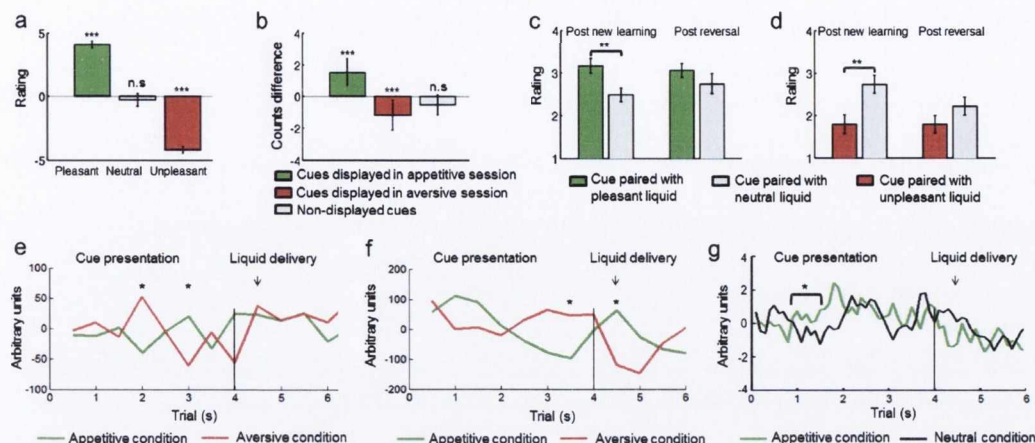


Figure 6.3. Behavioural results. a, Ratings for the pleasant, neutral and unpleasant liquids (-5 being very unpleasant and 5 very pleasant). *** indicates a significance of $p < 0.001$ as computed by one sample t-tests comparing the mean of the different liquids against a mean of 0, n.s. stands for not significant. b, Difference in the number of times a cue is preferred after - before the experiment. *** indicates a significance of $p < 0.001$ as computed by one sample t-tests comparing the mean of the different liquids against a mean of 0. c, d Ratings for the cue paired with the pleasant (c) or unpleasant (d) liquid and the cue paired with the neutral liquid after a few trials after a new pair of cue has been presented (Post new learning) and a few trials after a reversal has occurred (Post reversal). A rating of 1 indicates that participants strongly dislike the cue whereas a rating of 4 indicates that they strongly like it. ** indicates a significance of $p < 0.01$ as computed by two sample t-tests comparing the means of the ratings for the cues paired with pleasant/unpleasant and neutral liquids. e-g, Conditioned responses. e, f, Time course for subjects' pulse (e) and respiration (f) averaged across all trials in the appetitive (green line) and aversive (red line) conditions. * indicates a significance of $p < 0.05$ as computed by two sample t-tests comparing the means of the amplitude for the appetitive and aversive conditions. g, Time course for pupil diameter in response to cues paired with the pleasant liquid (green line) and the neutral liquid (black line) averaged across all trials in the appetitive session for the 10 subjects showing reliable amplitude in their pupil diameter. A one-tailed paired t-test for a time window 0.8-1.5s revealed a significant decrease in constriction when participants were presented with cues paired with the pleasant liquid ($p < 0.05$).

6.3.2 fMRI results

We report results from our analyses within the amygdala using a height threshold of $p < 0.005$, with an extent threshold significant at $p < 0.05$ corrected for multiple comparisons. We first report results from our model-based learning algorithm (the HMM model). Next we report results from the Rescorla-Wagner and Pearce-Hall models (two instances of a model-free learning algorithm).

6.3.2.1 Model-based learning algorithm: HMM model

-Expected value signals

We first investigated BOLD activity in the amygdala correlating with expected value (EV) signals at the time of cue presentation. In the appetitive session, we found significant activity positively correlating with expected value in the medial part of the right amygdala, corresponding to the basolateral complex (Figure 6.4a in green, MNI [x y z] [10 -10 -18], $T = 6.29$, $k=28$ voxels). In the aversive session, activity positively correlating with expected value was found in the centromedial complex of the left amygdala (Figure 6.4a in red, [x y z] [-27 -2 -9], $T = 5.63$, $k=44$ voxels; [x y z] [-17 -15 -14], $T = 5.41$, $k=69$ voxels), such that the greater the activity in these areas, the less an aversive outcome is predicted to occur. We also looked for areas correlating negatively with EV in both the appetitive and aversive sessions, that is, areas showing an increase in activity the less a positive outcome was predicted to occur given the cue. We did not find evidence for such activity in the amygdala in either the appetitive or the aversive session at our statistical threshold.

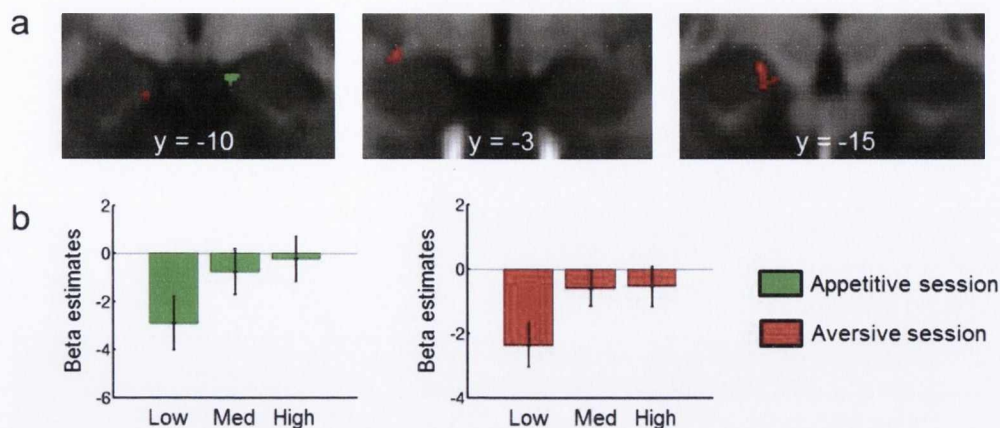


Figure 6.4. Expected value signals from the model-based learning algorithm model in the amygdala. a, Blood oxygen level-dependent (BOLD) signals positively correlating with the magnitude of the expected value of the cue were found in the basolateral complex in the appetitive session (in green) and in the centromedial complex in the aversive session (in red). b, Plots showing the beta estimates for low, medium and high categories of expected rewards in the appetitive (green) and aversive (red) sessions in the clusters activated using the leave-one out method.

-Precision signals

Next, we examined amygdala activity correlating positively with precision or else correlating negatively with precision during both the appetitive and aversive sessions. While no significant negative correlation was found with precision, we did find significant correlations with precision signals during both the appetitive and aversive sessions within our centromedial complex ROI (appetitive session: $[x\ y\ z]$ [25 -1 -10], $T = 4.12$, $k=44$; aversive session: $[x\ y\ z]$ [27 -5 -10], $T = 5.31$, $k=115$; $[x\ y\ z]$ [18 -2 -16], $T = 4.75$, $k=44$) (Figure 6.5a). To test whether there was a significant overlap between these clusters in the appetitive and aversive sessions, we performed a formal conjunction analysis (at our omnibus threshold of $p < 0.005$ with a cluster extent of $p < 0.05$). In this

contrast we found a common area activated by precision signals in the appetitive and aversive sessions in the centromedial complex of the amygdala ($[x\ y\ z]\ [24\ -4\ -9]$, $T = 3.52$, $k=23$)(Figure 6.5c).

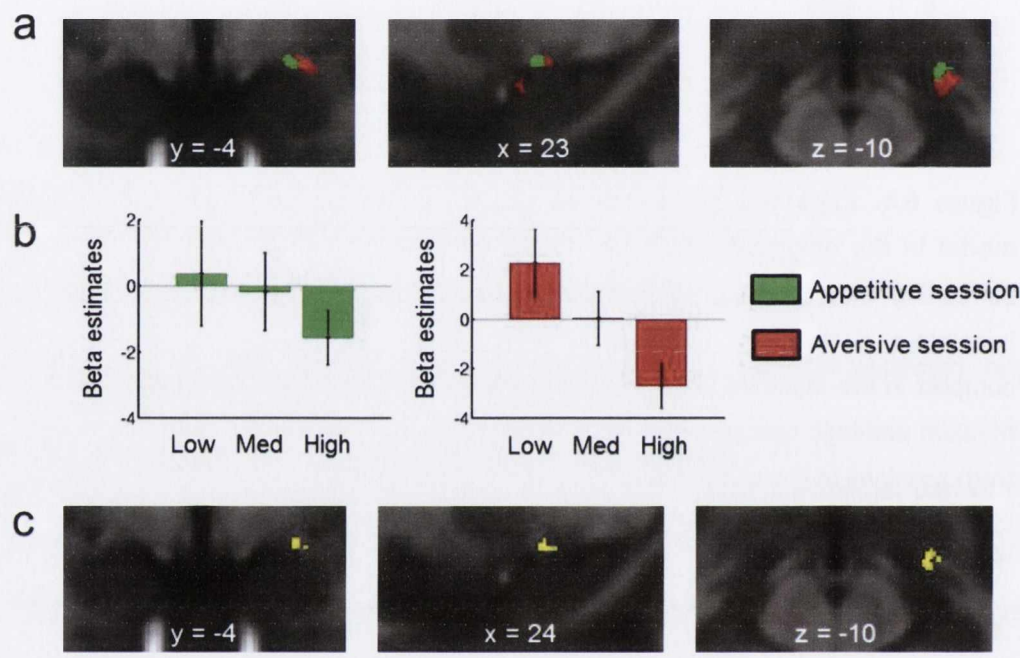


Figure 6.5. Precision signals from the model-based learning algorithm in the amygdala. a, Blood oxygen level-dependent (BOLD) signals correlating with the precision of the cue were found in the centromedial complex of the amygdala in both the appetitive session (in green) and aversive session (in red). b, Plots showing the beta estimates for low, medium and high categories of precision in the appetitive (green) and aversive (red) sessions in the clusters activated using the leave-one out method. c, Results from formal conjunction analysis of precision signals from the appetitive and aversive sessions in the centromedial complex.

6.3.2.2 Model-free learning algorithms

Although the behavioural fits of the model-free algorithms to the behavioural data were not better than that which was obtained by the random baseline model suggesting they are doing poorly at accounting for behaviour on the learning task, we still endeavored to determine the extent to which such model-free algorithms could account for BOLD activity in the amygdala. Because of the poor behaviour fits we could not generate reliable parameter estimates from the behaviour. Nevertheless, for the fMRI analysis we cycled through a range of plausible free parameters and fit the predicted time series generated by each of these free parameters separately to the fMRI data. This allowed us to cover the space of plausible free parameters, and therefore report any activations correlating with the predictions of the model-free algorithms in the amygdala. However, because of the absence of a reliable fit of these models to the behavioural data, and the fact that we had to run multiple (though correlated) analysis models in order to find effects, the following results should be interpreted with caution. Nevertheless, we report them for completeness:

-Rescorla-Wagner

Positive EV signals estimated using learning rates of 0.4 through to 0.6 were found in the aversive session in an area very close to the one activated by the same contrast in the HMM GLM (peaking at the learning rate of 0.5: [x y z] [-17 -15 -17], $T = 5.07$, $k=35$).

-Pearce-Hall

Using a free parameter of 0.1, we found some activity in the right amygdala negatively correlating with EV signals in the aversive session, i.e the greater the activity the more

the unpleasant tasting liquid was expected ([x y z] [-27 -6 -18], $T = 5.31$, $k=30$; [x y z] [28 -5 -23], $T = 4.49$, $k=37$). A positive correlation with EV was found in the left amygdala for the aversive session with free parameters of 0.2 through to 0.5, which was maximal at the free parameter of 0.4 ([x y z] [-19 -3 -17], $T = 5.02$, $k=51$). Positive correlations with EV were only found in the appetitive session using free parameters of 0.3 to 0.4 (maximal at free parameter of 0.3: [x y z] [-10 -6 -18], $T = 4.11$, $k=34$).

6.3.2.3 Model-comparison on BOLD data

In order to determine whether BOLD activity in the amygdala is better accounted for by the HMM than by the model-free learning algorithms, we performed a Bayesian Model Selection (BMS) analysis. The expected value contrasts from our “model-based” Hidden state Markov switching model (HMM) and the “model-free” Rescorla-Wagner (RW) and Pearce-Hall (PH) models were used to compare BOLD activity in the amygdala separately for the aversive and appetitive sessions. In this model comparison, we included voxels within a 4mm sphere centered on the peak voxel of the amygdala activity correlating with expected value signals for the ‘model-based’ HMM using the leave-one out method, thereby avoiding a non-independence bias in the voxel selection. Given that the model-free algorithms failed to fit our behavioural data better than a random model, model-free parameters could not be reliably estimated from the behavioural data. To enable a model comparison in any event, we used the full range of plausible free parameters, that is from 0.1 to 0.6 with increments of 0.1 and fit model timeseries corresponding to each of these free parameter values to the fMRI data. We found that the model-based HMM outperformed both model-free algorithms with free

parameters ranging from 0.1 to 0.6 with an exceedance probability of 0.89 for the aversive session and of 0.73 for the appetitive session.

We also performed a similar BMS to discriminate between our “model-based” HMM and a simpler version of this HMM which uses Bayesian updating but does so in a manner resembling a more “model-free” algorithm. The essential difference between these two HMMs is that the “model-based” HMM does not allow for a reversal without moving from a non-reversal state to a possible reversal state. Note that the expected reward signals generated by these two HMMs are highly correlated whereas the precision values are not. Hence, we compared neural activity within the contrasts showing activity positively correlating with precision signals including voxels within a 4mm sphere centered on the peak voxel of the amygdalar activity correlating with precision signals for the “model-based” HMM using the leave-one out method, thereby avoiding a non-independence bias in the voxel selection. We found that activity was better explained by precision signals estimated by the “model-based” HMM in both the aversive and appetitive sessions (Aversive session: exceedance probability=0.99; Appetitive session: exceedance probability=0.57).

6.4 DISCUSSION

In this study, we used a Pavlovian conditioning task with a rudimentary higher-order structure in both appetitive and aversive domains to investigate whether neural activity in the human amygdala reflects learning that requires access to model-based representations. By comparing neural activity correlating with expected value signals generated by model-based versus model-free learning algorithms using a Bayesian model selection (BMS) procedure, we have been able to show that in at least some

parts of the human amygdala, activity during Pavlovian conditioning is better accounted for by a model-based rather than a model-free algorithm.

One of the critical distinctions between the model-free and model-based learning algorithms in the present study is that while the expected value of a stimulus previously paired with the unpleasant outcome is still low following reversal of contingencies because that was the value it had before reversal in a model-free system, the expected value of this stimulus will become high in a model-based system because it incorporates the knowledge that after a reversal stimulus values switch (i.e. there is full resolution of uncertainty when a reversal occurs). We have captured model-based representations in formal terms using an elementary Bayesian Hidden Markov computational model that incorporates the task structure (by encoding the inverse relationship between the cues and featuring a known probability that the contingencies will reverse).

Our behavioural analysis demonstrated that participants showed evidence of conditioned responses to the conditioned stimuli and thus successfully learnt the associations between the different cues and outcomes. In a trial-by-trial analysis in which we correlated reaction times against the model predictions, we found that the HMM model predicted changes in reaction times over time as a function of learning better than the model-free alternatives, and that indeed the model-free algorithms did not predict variation in reaction times significantly better than chance.

In the imaging data, we found trial-by-trial positive correlations of model-based expected values in an area consistent with the basolateral complex of the amygdala according to the Mai atlas in the appetitive session, and in areas in the likely vicinity of the centromedial complex in the aversive session (Mai et al., 2008). It is interesting to

note that activity in these same areas (i.e basolateral versus centromedial complex) has been found to correlate with expected value signals generated by a simple RW model in a recent reward versus avoidance instrumental learning task (in an appetitive versus aversive context respectively) (Prevost et al., 2011). Using a BMS procedure, we found that amygdala activity correlating with expected value was best explained by model-based than model-free learning algorithms. Whereas the model-free system has received considerable attention in the past (O'Doherty et al., 2007), the more sophisticated and flexible model-based system, has been more sparsely studied particularly in relation to its role in Pavlovian learning. Thus, our results point to the need for integrating model-based representations and their rich adaptability into our understanding of Pavlovian conditioning in general, and of the role of the amygdala in implementing this learning process in particular.

Another important feature of the model-based algorithm featured in this study, is that as well as keeping track of expected value, this model also keeps track of the degree of precision in the prediction of expected value over the course of learning. This precision starts off low at the beginning of a learning session with a new stimulus because the expected value computation is very uncertain at this juncture, but once outcomes are experienced in response to specific cues, the precision in the estimate quickly increases. However this precision lessens again as the trial progresses because a reversal in the contingencies is increasingly expected to occur (hence the expected value becomes more and more uncertain). Signals correlating with precision were found to be located in the vicinity of the centromedial complex in both the appetitive and aversive sessions. Precision signals might play an important role in the directing of attentional resources toward stimuli in the environment. The presence of a precision signal in the centromedial amygdala in the present paradigm, could be a key

computational signal underpinning the putative role of this structure in directing attention and orienting toward affectively significant stimuli.

The presence of a precision-related signal in the amygdala during Pavlovian conditioning may relate to other findings in which the amygdala has been suggested to play a role in “associability” as implemented in a model-free algorithm such as the Pearce-Hall learning rule (Li et al., 2011; Roesch et al., 2010). Associability as defined in such a model is essentially a model free computation of uncertainty, the inverse of precision: associability is maximal when the absolute value difference between expected and actual rewards is greatest. However, in our case, an associability signal is clearly distinct from the signal we observe in the amygdala in the centromedial complex (even leaving aside the fact the signal we found is negatively as opposed to positively correlated with uncertainty). First of all, because the signal in our HMM is model-based, it changes to reflect anticipated changes in task structure (such as a reversal), whereas Pearce-Hall associability does not change to reflect anticipated changes in task structure, both rather changes only reflexively once contingencies have reversed. The model-based nature of our signal was confirmed by comparing the precision signal generated by the model-based algorithm with that generated by a model-free version of our Hidden-Markov Model: activity in the amygdala was best accounted for by the precision signal generated by the model-based algorithm. Finally, we checked the correlation between the precision signal we found here and an associability signal generated by the Pearce-Hall learning rule, and we found the correlation between these signals to be essentially negligible (with r ranging from -0.06 to -0.14), as opposed to being strongly negatively or positively correlated as would be anticipated were these signals to tap similar underlying processes.

The fact that in the present study we found model-based signals in the amygdala does indicate that this structure is capable of performing model-based inference even during Pavlovian conditioning. However, it is important to note that the findings of the present study do not rule out a role for this structure in model-free computations during Pavlovian conditioning. Indeed, while the model-free learning rules we used did not work very well in accounting for behaviour on the task (as indexed by changes in reaction times), we did find some evidence (albeit weakly) of model-free value signals in the amygdala as generated by either a Rescorla-Wagner or Pearce-Hall learning rule. Indeed, while using our HMM model we did not find evidence for aversive-going expected value signals in the aversive session (i.e. by showing an increase in activity the more the unpleasant tasting liquid was expected), we did find such a signal correlating with expected value as computed by a Pearce-Hall learning rule. As a consequence, we cannot rule out a contribution for the amygdala in model-free computations. It is important to note however, that in many tasks in which neuronal activity was found in the amygdala to correlate with the predictions of model-free learning algorithms (Elliott et al., 2004; Hampton et al., 2007; Prevost et al., 2011; Yacubian et al., 2006), such tasks were either not set up to discriminate the predictions of model-free versus model-based learning rules, or else the relevant model comparisons were not performed. Thus, it is entirely feasible that many of the computations found in the amygdala in previous studies correspond more closely to model-based as opposed to model-free learning signals. More generally, if indeed, both model-based and model-free signals are present in the amygdala during Pavlovian conditioning, then an important question for future research will be to address how and when these signals interact with each other.

To conclude, we have found in the present study evidence for the existence of model-based learning signals in the human amygdala during performance of a Pavlovian conditioning task with a simple task structure. These findings provide an important new perspective into the functions of the amygdala by suggesting that this structure may participate in model-based computations in which abstract knowledge of the structure of the world is taken into account when computing signals leading to the elicitation of Pavlovian conditioned responses. The findings also resonate with an emerging theme in the neurobiology of reinforcement learning whereby value signals are suggested to be computed via two mechanisms: a model-based and a model-free approach (Dayan and Daw, 2008; Doya et al., 2002). Whereas up to now, theoretical and experimental work on this distinction has tended to be focused on the domain of instrumental conditioning (Daw et al., 2005; Glascher et al., 2010; Hampton et al., 2006), the present study illustrates how similar principles may well apply even at the level of Pavlovian conditioning. Thus the distinction between model-based and model-free learning systems may apply at a much more general level across multiple types of associative learning in the brain. Furthermore, the present results provide evidence that model-based computations may be present not only in prefrontal cortex and striatum, but also in other brain structures such as the amygdala.

In this last chapter, the findings of the different studies presented in this thesis will be reviewed and considered in the light of their implication for understanding the role of human amygdala's main complexes (i.e., BLA and CeN) in reward processing and associative learning. Limitations, challenges and future directions for high-resolution fMRI of the amygdala will then be discussed.

7.1 ELUCIDATING THE FUNCTIONS OF HUMAN AMYGDALAR SUBREGIONS

This thesis has presented a body of research assessing the role of human amygdalar subregions in reward processing and associative learning. The development of a high-resolution imaging protocol (including a decreased voxel size, the unwarping of the functional data, a specific normalization of the amygdala, the forgoing of the application of extrinsic spatial smoothing and the removal of physiological fluctuations) specifically optimized for the amygdala has allowed us to considerably increase the spatial resolution of our data and extract reliable signal from the human amygdala's main complexes, namely the BLA and CeN. The works presented here highlight three main findings. First, BLA and CeN make different contributions to the Pavlovian-Instrumental transfer phenomenon, with the former being involved in its specific form and the latter being implicated in its general form. Second, BLA and CeN are differentially recruited in appetitive versus aversive contexts, with the BLA mediating appetitive signals and CeN mediating aversive ones. Finally, amygdalar subregions

represent various computational signals required for learning and evidence points to the fact that their activity is better accounted for by a model-based approach that incorporates a rich structure of the environment than a model-free approach that does not incorporate such knowledge. Taken together, these findings suggest that CeN may underly simple conditioned motivational influences on behaviour, while BLA provides a more complex representational role in behaviour. Furthermore, these results show that not only has hr-fMRI allowed us to replicate findings from rodent lesion studies, but has also allowed us to go beyond what can be achieved in animal experiments by pinpointing very specific and sophisticated computational signals in the human amygdala.

7.1.1 Summary of the thesis findings

Before discussing the main implications of the findings presented in this thesis, let us first briefly summarize the main conclusions of each of these studies.

In chapter 3, using a novel free operant task designed to elicit effects of both general and specific Pavlovian-Instrumental transfer (PIT), we have provided the first behavioural evidence of general appetitive Pavlovian instrumental transfer in humans. Making use of our high-resolution imaging protocol, we further showed the involvement of a region of human ventral amygdala within the boundaries of the BLA in mediating specific PIT, and that of the dorsal amygdala within the boundaries of the CeN in mediating general PIT effects in humans. This completely novel result in humans dovetails with the animal literature, which has shown that lesions of the BLA impaired specific PIT while lesions of the CeN impaired general PIT (Corbit and Balleine, 2005).

In chapter 4, we used a simple appetitive versus aversive Pavlovian task in combination with our hr-fMRI protocol specifically optimized for the amygdala to show for the first time a differential involvement of BLA and CeN in the learning of general appetitive and aversive contexts respectively. Despite the fact that such dissociation between amygdalar subregions in learning about positive versus negative contexts has been hinted at in both the animal (Holland et al, 2002; Petrovich et al, 2009) and human (Prevost et al, 2011) literatures, this hypothesis had never been formally tested before and represents an important novel finding.

In chapter 5, through the application of our high-resolution fMRI protocol along with a computational approach, we revealed partly distinct functional contributions of BLA and CeN in the representation of specific computational signals and as a function of the type of instrumental conditioning involved (reward versus avoidance learning) using a probabilistic reversal learning task (Prevost et al., 2011). Going beyond previous fMRI studies, the use of hr-fMRI allowed us establish the extent to which these signals are localized to specific subregions of the amygdala, both as a function of the type of computational signal involved (action reward value, prediction error and estimation uncertainty), and as a function of the type of learning involved (reward or avoidance learning). Furthermore, these findings highlight the fact that similar computational processes in these regions operate on reward and avoidance instrumental learning. Collectively, these results suggest that BLA and CeN appear to be differentially involved in reward and avoidance learning respectively, and provide unique insights into the computations being implemented within fine-grained amygdala circuits.

Finally, in chapter 6, we used a novel Pavlovian conditioning task in combination with our hr-fMRI protocol and sophisticated computational analyses to show that the

human amygdala, which is known to be a core structure for Pavlovian learning, contains neural signals consistent with model-based computations. These findings provide unique insights into the nature of encoding of valuation and learning signals in the human amygdala and furthermore, demonstrate that model-based processing is not unique to instrumental conditioning but is also implemented in the more fundamental (and hitherto assumed to be computationally simpler) type of learning considered here.

7.1.2 Specific versus general processes in the amygdala

In instrumental conditioning, animals learn the relationship between their actions and the outcome of those actions. Instrumental performance can be strongly influenced by incentive processes such as Pavlovian incentive learning, which mediates the excitatory effects of stimuli associated with outcomes (Lovibond, 1983). A paradigm well suited to examine the interaction of these incentive processes is Pavlovian-Instrumental Transfer or PIT (see Chapter 3 for a study of PIT). However, note that because a stimulus may produce 1) a general enhancement in responding as a result of any generally arousing or motivational effects that the stimulus has acquired through its association with reward generally (general PIT) or 2) an outcome-specific enhancement of a particular response as a result of a specific relationship with the unique sensory properties of a particular outcome (specific PIT), this paradigm also lends itself perfectly to study specific versus general processes in the brain. One of the main findings of this thesis is that while BLA seems to be involved in specific processes, CeN seems to be involved in general processes, concurrent with rodent lesion data showing a double dissociation of BLA and CeN in mediating specific versus general PIT (Corbit and Balleine, 2005). These results have several implications. First, they suggest that the

BLA is involved in the formation of Pavlovian incentives involving the association of a stimulus with the specific sensory features of the outcome and thus has a direct role in consummatory conditioning (which can be reflected by changes in chewing, licking, blinking and flinching). Second, the CeN mediates the influence of the general affective and reinforcing properties of these stimuli, and thus is more involved in preparatory conditioning (reflected by changes in heart rate, blood pressure, approach and withdrawal). Taken together, these results provide evidence for a structural basis for the distinct associations and processes involved in consummatory and preparatory conditioning first proposed by Konorski (Konorski, 1967).

7.1.3 Differential contribution of BLA and CeN in appetitive versus aversive representations

Another important finding of this thesis is that BLA and CeN seem to be differentially involved in the encoding of appetitive and aversive contexts respectively. Even though there are hints for such double dissociation in the animal literature, this has never been formally tested to our knowledge. In chapter 4, we provide evidence for a double dissociation of the roles of BLA and CeN in the encoding of appetitive (learning about the positive consequences of new Pavlovian stimuli) and aversive (learning about the negative consequences of new Pavlovian stimuli) respectively. These results concord well with rodent lesion data providing evidence for a double dissociation between BLA and CeN when learning about the influence of Pavlovian cues in a positive context (leading to food consumption) versus in an aversive context (leading to footshock) (Holland et al, 2002; Petrovich et al, 2009). They are also in agreement with our 'estimation uncertainty' results (reported in Chapter 5) which

revealed a double dissociation between BLA and CeN in a reward (leading to a positive outcome) versus avoidance (leading to a negative outcome) condition and which corresponded to an uncertainty metric having an exponential decay profile (similar to the exponential decays used in Chapter 4). This signal is consistent with the encoding of a positive versus aversive context as it was high at the beginning of each new session when participants had to encode the significance of each novel stimulus in the context of these positive versus aversive conditions and decreased over the course of learning (note that however, this study was not designed to specifically test this hypothesis and several alternative explanations could account for this signal, as discussed in section 4.4).

Note that this differential contribution of BLA and CeN in appetitive versus aversive processes is further corroborated by our expected reward (ER) value signals reported in Chapters 5 and 6, which reveal more activity in BLA in the representation of ER in appetitive conditions and more activity in CeN in the representation of ER in aversive conditions. Furthermore, the fact that this computational signal is implemented in these separate amygdalar areas both when applying model-free (Chapter 5) and model-based (Chapter 6) learning algorithms points to a general differential involvement of BLA and CeN in the representation of ER in appetitive versus aversive contexts respectively.

These findings may relate to the well-known anatomical projections of BLA and CeN. The BLA has independent projections to the ventral striatum and prefrontal cortex (especially the orbitofrontal cortex), connections that give it access to high-order response mechanisms (Everitt et al., 1999). In contrast, the CeN projects mainly to neuroendocrine and autonomic domains of the hypothalamus and brainstem, thus having a seemingly more basic function, highly conserved across species. It is also

known as the 'fight or flight' center of the brain, as it mediates reactions to potentially harmful stimuli and contains cells with responses to pain (Bernard et al., 1990; Kalin, 2004; Price and Amaral, 1981). Furthermore, from a neuroanatomical perspective, the BLA possesses a number of cortical like characteristics, consistent with the idea that it is phylogenetically more recent in the evolution of vertebrates whereas CeN represents a phylogenetically older group of nuclei (Alheid and Heimer, 1988; Swanson and Petrovich, 1998). As such, it could be argued that aversive processes may be mediated by the phylogenetically old and highly preserved CeN while appetitive processes may preferentially recruit BLA, which has evolved much more recently with the neocortex and is known to mediate more sophisticated processes, especially through its dense reciprocal connections with the prefrontal cortex. However, these differential contributions are unlikely to be exclusive as BLA and CeN are also highly interconnected and work together as a system, but they may provide a structural basis for the fact that aversive and appetitive stimuli elicit different types of conditioned responses.

7.1.4 The amygdala, a structure implementing model-based learning?

Demonstrating that amygdala activity was better accounted for by model-based than model-free learning algorithms (Chapter 6) is an important contribution of this thesis.

In a model-free learning system, state values are learnt without building an explicit model of the environment and consequently, without any estimate of the probabilities governing state transitions. Conversely, model-based learning algorithms use an internal representation of environmental contingencies in the form of a cognitive

map or model of the environment describing how different states of the world are connected to each other.

A long-held view concerning amygdala function is that it is involved in basic or primary functions. Indeed, research on amygdala has been prominently restricted to its role in detecting and reacting to threat and fear, most notably using Pavlovian fear conditioning paradigms (LeDoux, 1998, 2003; LeDoux, 2000a). However, research in recent years has revealed a much more elaborate picture of the amygdala, and evidence suggests that the amygdala has a very broad and abstract role, ranging from emotional learning and memory to complex social interactions (Adolphs, 2008, 2010; Baxter and Murray, 2002; Murray, 2007; Phelps, 2006). Furthermore, a computational account has started to emerge from reward learning studies that have revealed a central role of the amygdala in the computation of a number of learning signals, such as expected outcomes and prediction errors (Prevost et al, 2011; Elliott et al, 2004, Seymour et al, 2005; Yacubian et al, 2006; Hampton et al, 2007) (see also Chapter 5). Yet, these signals were estimated using standard model-free learning algorithms such as Rescorla-Wagner (Rescorla & Wagner, 1972) which fail to exploit higher-order structures such as interdependencies between different stimuli and subsequent outcomes. Indeed, whereas the model-free system has received considerable attention in the past, the more sophisticated and flexible model-based system, has been more sparsely studied. In chapter 6, we have captured model-based representations in formal terms using an elementary Bayesian Hidden Markov computational model that incorporates the task structure.

Therefore, our results have implications for understanding the role of the amygdala in performing inference on the basis of abstract states and point to the need of integrating model-based representations and their rich adaptability in future studies of

amygdala function. They also support an important role for model-based algorithms in describing the processes underpinning Pavlovian conditioning, suggesting that the distinction between model-based and model-free learning systems may apply at a much more general level across multiple types of associative learning in the brain. Finally, these findings imply that theories of amygdala function be modified to incorporate higher level functions such as model-based representations.

7.1.5 Reconciling our findings with the animal literature

The results reported in this thesis are not fully compatible with either theory of amygdala function described in the introduction chapter (see section 1.4). However, because 1) the 'parallel view' proposed by Balleine and Killcross (2006) relies heavily on their PIT findings (Corbit and Balleine, 2005), and 2) we have been able to replicate these findings of differential involvement of BLA and CeN in specific and general PIT respectively, it is reasonable to say that these results may support the 'parallel view', with the caveat that we can only provide indirect evidence for this claim. Note that however, hr-fMRI (and to a greater extent standard fMRI) is not powerful enough, at least in terms of temporal resolution, to discriminate between the 'serial' and 'parallel' views which would require a timescale of milliseconds as opposed to the several seconds afforded by fMRI. Nevertheless, despite the fact that 1) there is no study of aversive PIT reporting a double dissociation of BLA and CeN in mediating specific and general PIT to date and 2) the 'serial' view relies exclusively on studies of aversive conditioning, it has been suggested that available evidence points to substantial similarities in function between BLA and CeN in appetitive and aversive situations, with the BLA mediating outcome-specific associations of stimuli and the CeN mediating associations with the more general affective aspects of the outcome (Balleine and

Killcross, 2006). This statement is not consistent with our findings showing a differential involvement of BLA and CeN in appetitive and aversive situations (Chapters 4 and 5). Instead, our findings are consistent with the notion that these two areas may contribute differentially to learning in different contexts rather than both being engaged in a serial manner or indeed always being engaged in parallel, a finding broadly consistent with the observations of differential functions for BLA and CeN by Holland and Gallagher in recent lesion studies. Specifically, Holland et al., (2002) reported that BLA but not CeN was involved in mediating feeding potentiation due to the presence of a reward predictive Pavlovian stimulus, while Petrovich et al., (2009) reported that CeN but not BLA was involved in conditioned suppression of feeding due to the presentation of an aversive predicting conditioned stimulus. It is entirely feasible that the types of processes being recruited in those previous studies and ours present some similarities. Therefore, rather than a strictly serial processing structure, or indeed, a fully parallel one, these findings suggest that in the human brain as well as in the rodent, different components of the amygdala might be differentially recruited during different forms of learning between cues, actions and outcomes in different contexts.

Reconciling our findings of differential contributions of BLA and CeN in specific versus general processes on the one hand and that of appetitive versus aversive representations on the other hand may not look straightforward at first sight. However, an intuitive possibility would be that while aversive contexts tend to recruit general processes (eg: fight or flight reactions), appetitive contexts may be more likely to recruit sensory specific mechanisms (eg: "I am interested in this particular reward"). This idea concurs with a line of research which has suggested that it is clear that BLA and CeN are involved in associative processes underlying both appetitive and aversive emotional behaviour, and that the inconsistencies reported concerning their respective roles can

be resolved around the hypothesis that CeN and its associated circuitry underlies simple conditioned motivational influences on behaviour, while BLA provides a more complex representational role in emotionally charged decisions and voluntary behaviour (Everitt et al., 2000).

Note that, however, the animal literature is not clear-cut on this topic as numerous contradictory findings have been reported. Several explanations could account for this. First, most paradigms of aversive conditioning have used only one outcome, which only permits inferences about general processes (as specific processes require distinct outcomes with their own sensory features). Second, the animal literature is extremely vast and has reported a large number of contradictory findings over the past two decades, which makes it difficult to reconcile any human finding with rodent data, in particular lesion studies. This phenomenon may arise in part from the fact that the location of damage, its extent, and its timing often differ between studies aimed at lesioning the same nuclei, rendering comparisons unreliable. Furthermore, different training and testing protocols may affect differentially the outcome of experiments intended to determine the role of specific nuclei. Therefore, discrepancies concerning the role of BLA and CeN stem not only from the use of heterogeneous anatomical, cellular, and molecular tools to probe the role of these nuclei but also from different behavioural protocols that are concerned with different forms of behaviour and learning.

A fruitful avenue for future research that may help integrate and reconcile the results reported in the current thesis with those reported in the animal literature could be to examine specific and general PIT using both appetitive and aversive conditioning. With this 2*2 factorial design, it would be possible to directly compare specific and general processes as well as appetitive and aversive conditioning in BLA and CeN.

7.2 LIMITATIONS OF HR-fMRI

Despite the promise of our hr-fMRI protocol in furthering the understanding of amygdalar subregions in reward learning in humans, it has a number of limitations which must be closely attended to if this field is to continue to progress.

7.2.1 Limitations of our hr-fMRI protocol

7.2.1.1 Defining the amygdala

A few brain atlases are available to help to define the boundaries of what is termed the amygdala (Duvernoy, 2005; Mai et al., 2008). However, these rely on very few (old) brains, which may add variability to the exact size/location of the amygdala. Furthermore, the literature on the human amygdala has often referred to 'the extended amygdala', where foci have been reported in reference to the amygdala because of its anatomical and functional relationship to CeN, when they were actually in the basal forebrain. Note that data regarding the precise location of this extended amygdala is lacking and the spatial resolution of standard fMRI studies has made it difficult to identify whether activations were falling within this extended amygdala or neighboring sublenticular regions.

Therefore, a more accurate description of the amygdala boundaries is needed in order to determine the functional domains of the amygdala proper and 'the extended amygdala'.

Other issues related to amygdala delineation are 1) the time required to manually segment ROIs in both hemispheres, and 2) issues pertaining to observer bias, where all ROIs in a study should ideally be traced by a single observer to maintain consistency

across participants. An important future avenue of research could be automated segmentation since it would reduce both time requirements and observer biases.

7.2.1.2 Segmentation of amygdalar subregions

To date, there is no ideal way to delineate amygdalar subregions. It is impossible to reliably identify different nuclei or even complexes on current high-resolution structural scans (typically 1 x 1 x 1 mm). Although it is possible to identify some boundaries between BLA and CeN at higher resolutions with sequences specifically adapted for the amygdala, these boundaries are only visible on a subset of slices, still preventing reliable segmentation across the whole amygdala.

The manual segmentation used in this thesis to delineate amygdalar subregions is far from perfect but nonetheless, the best compromise available to date. Two alternative methods have been proposed to anatomically delineate human amygdalar subregions using MRI: Amunts et al., (2005) used post-mortem cytoarchitectonics to develop a probabilistic map of amygdala nuclei, while Solano-Castiella et al., (2010) used diffusion-weighted MRIs to cluster amygdala subregions based on the predominant diffusion direction. However, although both of these approaches are promising future directions, neither of these segmentation methods are appropriate for the following reasons:

In the case of the probabilistic map of Amunts et al., the identified subregions extend far beyond the anatomical boundaries of the amygdala to such an extent that almost as much of the ROI is in adjacent temporal lobe white matter as is in the amygdala proper (see Figure 7.1a). This suggests a mismatch between the localization of the amygdala and that of the probabilistic map (Amunts et al., 2005). This may be

due to the fact that the post-mortem brains used to derive the map were from an old population (average age 65). Given that gross structural changes are known to occur in the brain with aging such as atrophication of brain structures and increase in ventricular size, differences in amygdala shape and location between young samples (standard for fMRI studies on healthy controls) and an older sample might account for such differences, rendering the Amunts map unsuitable for use in fMRI studies. Note that these probabilistic maps have been used in the literature, although very infrequently (Ball et al., 2007).

The Solano-Castiella et al. diffusion-based segmentation method yielded two distinct areas of the amygdala along a medial to lateral dimension (Solano-Castiella et al., 2010). Unfortunately, this delineation does not map clearly onto the known nuclear structure of the amygdala and the known location of the BLA and CeN complexes (which are better differentiated along a ventral to dorsal dimension). This might reflect the possibility that the diffusion-based segmentation method is likely more sensitive to myeloarchitecture as opposed to cytoarchitecture.

Our approach, which has been driven by delineating the approximate boundaries of BLA and CeN using a detailed anatomical atlas of the amygdala (Mai et al., 2008) does not rely on anatomical grounds, and as such, can be seen as controversial. In addition, delineating these complexes on our template does not account for inter-individual variability in the location of these subregions. This notwithstanding, it appears that there is a close correspondence between the delineation used in our studies on our specifically normalized amygdala template, and the localization of the nuclei from the Mai atlas, as shown in Figure 7.1b and c. This suggests that we have captured, to a reasonable level of approximation, the location of the boundaries between these two substructures (note that the probabilistic maps of

Amunts and colleagues are not a close correspondence to the Mai atlas, as shown on Figure 7.1a,).

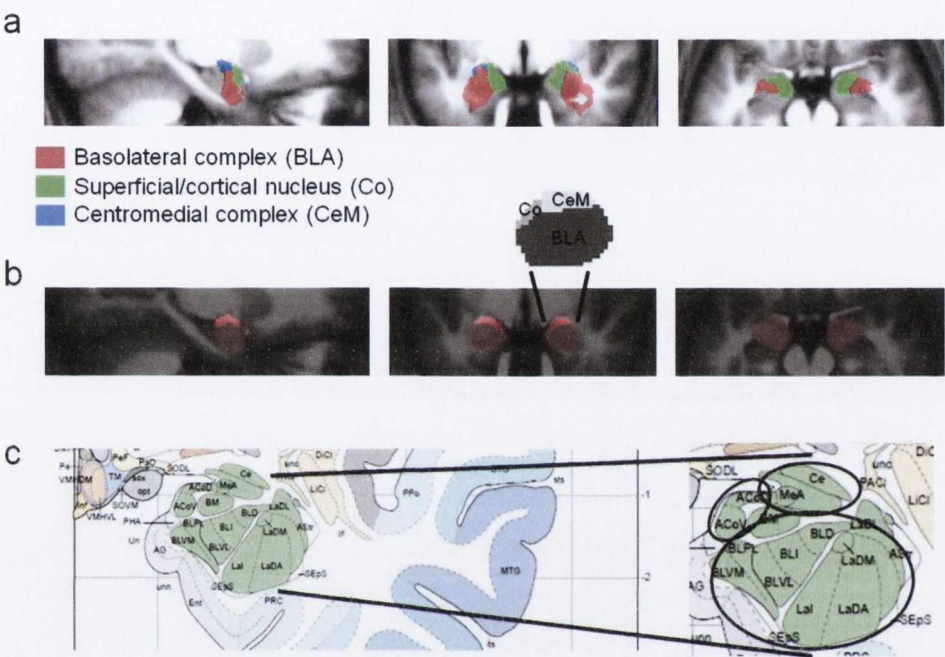


Figure 7.1. Amygdalar subregions according to Amunts' probabilistic maps (a), our template (b) and the Mai atlas (c). In a and b, sagittal (left), coronal (middle) and axial (right) sections are shown. In c, a coronal section is shown.

Nevertheless, two recent studies describe promising methods of segmenting the amygdala. Bach et al. (2011) used diffusion weighted imaging with probabilistic tractography to reveal deep and superficial amygdala nuclear groups projections, and a potentially powerful approach for in vivo amygdala parcellation. However, although the dissociation is along a dorso-ventral dimension which is in agreement with the location of BLA and CeN (as compared to the medio-lateral dimension shown in Solano-

Castiella et al.), it is still very coarse and not a good approximation of the location of these two complexes of the human amygdala according to the Mai atlas (Bach et al., 2011).

Finally, the most promising avenue so far has been proposed by Saygin et al (2011) who developed a method to segment the human amygdala into its four major nuclei using only diffusion-weighted imaging and connectivity patterns derived mainly from animal studies (referred to as Tractography-based Segmentation). This connectivity-based segmentation of human amygdala nuclei using probabilistic tractography is the best estimation available to date and a probabilistic map toolbox is to be released (Saygin et al., 2011). Once this tool is available, anatomical ROIs for the four main nuclear groups (including the BLA and CeN) will allow researchers to use independent anatomical ROIs for their ROI analyses when looking for dissociations of function among different amygdalar subregions.

In conclusion, although there is no perfect way of segmenting amygdalar subregions to date, our method seems to be a good approximation despite its limitations, but new methods such as the one just mentioned are very promising for the future of amygdala segmentation.

7.2.1.3 Coupling model-based fMRI with hr-fMRI of the amygdala

An important trade-off needs to be made when collecting high-resolution images of the amygdala because only a limited number of slices can be acquired without exceeding a reasonable acquisition time per volume (repetition times are usually kept around 2s). This trade-off between volume coverage, voxel size, temporal resolution and size of the datasets acquired is very likely to result in slice prescriptions that

exclude a large number of brain areas outside your particular region of interest and its immediate surroundings. An issue associated with this restricted coverage is that it prevents exploration of the functional connectivity between your ROI and areas outside your slice prescription (see section 7.3.2 for more details). Furthermore, despite allowing us to test specific computational hypotheses, coupling hr-fMRI and computational modeling, will inevitably limit the possibility of detecting unexpected findings that differ from a specific a priori hypothesis of a very specific signal in a very specific brain region. To take these issues into account, it is therefore necessary not to narrow research by 1) replicating hr-fMRI experiments at standard resolution to assess how areas of the brain known to have strong structural connectivity such as the prefrontal cortex and striatum (see section 1.3) interact with the amygdala and 2) continue to employ conventional trial-based analyses alongside the model-based approach.

7.2.2 The amygdala: a polynuclear structure

As mentioned throughout Chapter 1, the amygdala is comprised of many more than two divisions, which should not be ignored. The 13 nuclei composing the amygdala can further be congregated in 6 main nuclei: the lateral, basolateral and basomedial nuclei form the basolateral complex (BLA), the central and medial nuclei form the centromedial complex (CeN) and the cortical nucleus (Co). To date, the animal literature has principally focused on the basolateral and central nuclei which are the homologous of the BLA and CeN in humans. However, to gain a thorough understanding of amygdala function, the roles of other nuclei of both the rodent and

human amygdala should also be explored. Although this should be feasible to some extent in rodents, it remains a major limitation of current hr-fMRI methodology.

As touched on above, despite the gains in resolution afforded by high-resolution imaging, anatomically discriminating BLA from CeN is difficult due to the relative absence of anatomical landmarks on structural scans. Even though recent advances in hr-fMRI sequences of structural imaging allow for segmentation of BLA and CeN in a small number of slices, further gain in resolution is required for these complexes to be reliably differentiated. Therefore, it is needless to say that going beyond differentiating these two regions is still problematic. However, with the increasing availability of 7T scanners, there is hope that higher magnetic fields will ultimately achieve the resolution required to investigate more than two large divisions of the amygdala, with the caveat of increased signal dropout and distortion in anterior and medial amygdala (Krasnow et al., 2003).

7.2.3 Limitations inherent to hr-fMRI (and fMRI)

Most of the limitations that apply to fMRI also apply to hr-fMRI, because the methodology relies on similar grounds (also see section 1.5.4 for issues associated with fMRI). One of the most prominent limitations of this methodology is that fMRI is correlative in nature, meaning that it is impossible to establish causal links between activity in a particular brain region and subsequent behaviour; thus this activity may simply be epiphenomenal.

Even though hr-fMRI has been developed to enhance spatial resolution of fMRI data, this resolution is nonetheless insufficient to provide insight into fine-grained coding properties conveyed at the single neuron level. The BOLD signal expressed in an hr-

fMRI voxel is thought to reflect neural processing from several thousands of neurons rather than spiking activity of single neurons (Angenstein et al., 2009; Logothetis et al., 2001; Logothetis and Wandell, 2004). As a result, relating hr-fMRI amygdalar subregion findings to single or multi-unit recordings in animals is obviously extremely challenging, and such comparison may thus be better suited with local field potentials (LFPs) or other measures of the activity of populations of neurons.

Another limitation of hr-fMRI is the modest temporal resolution. Hr-fMRI focuses exclusively on spatial resolution and usually has a similar temporal resolution to that of standard fMRI. As explained in section 1.5.2, fMRI is an indirect measure of neuronal activity in which the BOLD signal peaks several seconds after neural response onset. Consequently, hypotheses requiring knowledge of millisecond timing differences is not achievable using hr-fMRI alone. A promising avenue would be to combine the high spatial resolution of hr-fMRI structural and functional data with the high spatial and temporal resolution of intracranial recordings in human patients suffering pharmacologically resistant temporal lobe epilepsy. Although such patients are not widely available, extant data suggests that multiple contacts in the amygdala (and throughout the MTL) can be localized to specific subregions using hr-fMRI. Moreover, integrating hr-fMRI, single unit and LFP data from these patients will not only advance understanding of amygdala function, but will also further the understanding of the relationship between BOLD activity in the amygdala and direct neuronal activity (Ekstrom et al., 2009). More details about the strong potential of this coupling can be found in section 7.3.1.

7.3 CHALLENGES AND FUTURE DIRECTIONS FOR HR-fMRI

7.3.1 Combining hr-fMRI with other approaches

On account of these limitations, it appears clear that imaging the amygdala with hr-fMRI is extremely challenging. For this reason, combining the results from hr-fMRI studies with data derived from other approaches will be an important direction for future research. Imaging methods that afford greater temporal resolution such as magnetoencephalography (MEG) or electroencephalography (EEG) will be valuable when studying cortical areas (unfortunately, these techniques are not designed for small subcortical structure deeply buried in the brain such as the amygdala, hence extracting decent signal for this area can prove difficult). To be able to draw causal inferences about the necessity of a particular brain structure for a particular function, one needs to investigate the effects of disruption of this structure either by studying patients with a permanent lesion to this region or by inducing temporary inactivation in this region through transcranial magnetic stimulation (TMS). However, both of these techniques have critical limitations when it comes to studying the amygdala. First of all, TMS is still in its infancy and concerns such as optimal frequency and duration of stimulation remain open issues (Machii et al., 2006). But, more importantly for our considerations, although it may be feasible to temporarily inactivate the amygdala with TMS as surface coil technology improves to target deeper structures, it is unlikely to be selectively inactivated due to its small size. As for lesions, a prominent methodological problem with amygdala-lesioned patients is the rarity of observing complete amygdala damage without observing damage to the adjacent cortex. Even selective lesions to the amygdala are likely to affect neighboring regions due to the presence of fibers of passage, and the disruption of these fibers is a major issue (Dupont et al., 2001; Goulet et al., 1998). However, patients with Urbach-Wiethe syndrome provide a unique

exception to these issues due to the relative selectivity of the calcification of the amygdala (at least until a certain age) (Markowitsch et al., 1994; Newton et al., 1971; Tranel and Hyman, 1990), but even with these patients there may be some calcification of the fibers of passage, and the lesions are not specific to a particular amygdalar subregion. In addition, the extreme rarity of such patients makes it difficult to accumulate even small samples of such patients.

A technique that may be much more successful in combination with hr-fMRI is intracranial electroencephalography because it can directly measure cellular activity from the amygdala and has much fewer limitations. As part of their presurgical screening, patients with epilepsy have depth electrodes implanted that can record from single cells at various locations along the electrodes, among them the amygdala (Fried et al., 2002). These recordings can be isolated to specific areas of the amygdala when combined with structural hr-fMRI, and have the potential to elucidate whether human amygdala cells have similar properties to that of animals as well as to determine unique aspects of the function of human amygdala cells, especially when it comes to more complex behaviours probably unique to humans, such as complex social interactions. Therefore, correlations between hr-fMRI measurements and electrophysiological measurements using depth electrodes is an exciting avenue for future research on amygdalar subregions.

Note that for continued progress in this field, it will also be crucial to compare hr-fMRI results with findings from the animal literature, be they from lesion or single or multi-unit neurophysiology studies.

7.3.2 Investigating the interaction between amygdalar subregions and other brain structures

Before being able to investigate the separate roles of different human amygdala nuclei and the interactions between amygdalar subregions and the rest of the brain, it will be necessary to further improve the spatial resolution of hr-fMRI so that amygdalar nuclei can be routinely and easily segmented. Therefore, an obvious route for future research will be to improve hr-fMRI techniques by developing new sequences where the voxel size is decreased further without compromising the SNR so that amygdalar nuclei can be easily delineated and functionally studied. Another important avenue will be to develop tools allowing very accurate segmentation of amygdalar nuclei so that these small areas can be coregistered across participants and permit to draw inferences at the group level.

In order to capitalize on enhanced spatial resolution, an important avenue of future research will exploit techniques investigating functional connectivity between amygdalar subregions and other cortical and subcortical brain regions associated with associative learning and reward processing. Descriptions of the circuitry and functions of the amygdala treat it as a gateway with a set of sensory inputs and outputs to effector sites; yet, a consideration of reentrant pathways and interactions between the amygdala and the rest of the brain is lost in this model. However, the OFC and the prefrontal cortex in general, which are strongly interconnected with different amygdalar subregions in both rodents and primates (Krettek and Price, 1977; McDonald, 1991a), have received some attention regarding their interaction with these subregions. For instance, some rodent studies combining lesions and electrophysiology have shown that interconnections between BLA and OFC afforded access to a further level of control over goal-directed behaviour (Gallagher and Schoenbaum, 1999; Hatfield et al.,

1996; Schoenbaum et al., 1998, 1999). Hr-fMRI could also be used to investigate the effects of amygdala lesions on processing in other brain regions and vice versa in humans (Morris et al., 2001b), which would provide a direct test of the manner in which the amygdala influences and is influenced by other areas. In a reward-related standard fMRI study, Hampton et al. (2007) scanned two amygdala-lesioned patients to elegantly reveal the absence of vmPFC activity in these patients in reversal learning as opposed to their matched controls who exhibited a strong activation in this area (Hampton et al., 2007). Similar studies aimed at addressing such influences using hr-fMRI will be essential if we are to gain a complete picture of amygdala function in reward learning in humans, although as explained in Section 7.3.1, an important caveat is that patients with discrete lesions to specific subregions of the amygdala are extremely difficult to find (therefore, an alternative strategy may be to investigate BOLD signal in amygdalar subregions in patients with OFC lesions for instance).

Another important pathway for amygdala dependent learning is through its innervation of the striatum, thought to be at the center of an interface between limbic processing and motor output. While the anterior division of the BLA provides a widespread innervation of both the caudate nucleus and nucleus accumbens (Groenewegen et al., 1990; Kelley et al., 1982; McDonald, 1991b), the posterior BLA innervates the more medial portion of the nucleus accumbens in particular (Groenewegen et al., 1990). Interestingly, in concordance with our PIT study and that of Corbit et al. (2005) where BLA was shown to be involved in specific PIT while CeN was involved in general PIT (Corbit and Balleine, 2005), rodent lesion studies indicate that while the shell of the nucleus accumbens (NAc) mediates specific PIT, the core of NAc mediates general PIT (Corbit and Balleine, 2011; Corbit et al., 2001; Hall et al., 2001), supporting the suggestion that a functional circuit exists between the amygdala and

NAc. This raises interesting avenues for future research on PIT and other amygdala-based learning paradigms, such as designing experiments optimized for both dynamic causal modeling and high-resolution imaging to assess the functional link between BLA and NAc shell on the one hand, and CeN and NAc core on the other hand.

To summarize, amygdala research in humans will greatly benefit from greater consideration of functional connectivity in future hr-fMRI studies. Hr-fMRI possesses a unique ability to non-invasively measure activity in multiple areas simultaneously and has the potential to provide insight into the functional connectivity between the amygdala and brain areas it is strongly structurally connected to, such as the prefrontal cortex and striatum (which can be easily covered in the reduced field of view of hr-fMRI sequences adapted for the amygdala due to their respective locations). Although a few studies have explored this potential using standard fMRI (Hampton et al., 2007; lidaka et al., 2001; Morris et al., 1998; Zald et al., 1998), it has yet to be widely capitalized upon.

7.3.3 Application to the human clinical domain

Not surprisingly given its scope of influence, the amygdala is involved in a number of emotional disorders such as anxiety and depression as well as in more pervasive disorders such as autism and schizophrenia. Ultimately, hr-fMRI could also offer an exciting means of evaluating amygdalar nuclei dysfunction in pathologies affecting amygdala function in these disorders, and inform neurocircuitry models of such disorders as well as assist in the diagnosis and the prediction of treatment response.

Of particular interest to the field of associative learning and amygdala research is an anxiety disorder called Post-Traumatic Stress Disorder (or PTSD). This disorder is

characterized by an impaired ability to learn that cues previously associated with danger no longer represent a threat. This impairment can easily be measured using fear extinction paradigms which test for a decrease in conditioned fear responses that normally occur when a conditioned stimulus (CS) is repeatedly presented in the absence of the aversive unconditioned stimulus (US). Consistent with an interpretation of PTSD as a syndrome of deficient extinction ability, one current model of PTSD which has been highly influenced by knowledge from rodent fear conditioning research given the phenomenological parallels between the two, argues that PTSD is characterized by hyperarousal of the amygdala and insufficient top-down control by the medial prefrontal cortex and the hippocampus particularly during extinction (Milad et al., 2009; Rauch et al., 2006). More specifically, it has been theorized that the vmPFC exerts inhibition on the amygdala, and that a defect in this inhibition could account for the symptoms of PTSD (Koenigs and Grafman, 2009). Notably, it has been shown that fear extinction in rats was associated with increased levels of synaptic inhibition in fear output neurons of CeN (Amano et al., 2010). Although this is consistent with our findings of aversive processes being represented in CeN, an important future avenue will be to develop new fear extinction paradigms in humans to test whether fear extinction takes place in the human CeN, which may eventually lead to improved extinction-based behavioral therapies for anxiety disorders.

An interesting example of nuclear specificity of human amygdala damage in brain diseases that has already received some attention concerns patients with drug-refractory temporal lobe epilepsy, in whom the lateral nucleus and the basolateral nucleus are the most damaged regions, while the central nucleus remains rather intact (Pitkänen and Amaral, 1998). Interestingly, experimental animal models of status epilepticus have also shown that the damage is nucleus specific (Tuunanen et al.,

1996). There is therefore reason to believe that an understanding of the behavioural consequences of nucleus specific amygdala damage in human brain diseases will eventually be developed.

7.4 CONCLUDING COMMENTS

Despite the limitations of hr-fMRI and the challenges associated with studying this structure, the hr-fMRI studies of the amygdala carried out in this thesis represent a first small yet important step in the understanding of the functional organization of the human amygdala at the subregional level. The main findings to be highlighted from these experiments are that BLA and CeN seem to make different contributions to 1) specific versus general processes and 2) appetitive versus aversive processes. Furthermore, we provide important novel insight into human amygdala function by demonstrating its implication in model-based inferences during Pavlovian conditioning. Not only does this research coincide with animal data, but it also goes beyond what is achievable from animal studies alone, and provides a solid basis for refining how amygdala function has been conceptualized.

As the current thesis makes evident, hr-fMRI is the first widely accessible technique having the potential to dramatically contribute to our understanding of the amygdala as well as any other heterogeneous brain region and will unquestionably fulfill its promise to bridge the gap between fMRI data in humans and lesion and electrophysiological studies in animals. The works described here provide preliminary support for functional heterogeneity among human amygdalar subregions and will enable future studies to address specific hypotheses motivated by the animal literature

and the studies presented here. However, advancing knowledge at this fine-grained level will also require converging evidence across methodologies.

In conclusion, understanding of amygdala function in the past decade has advanced at an unprecedented rate, most likely due to the fact that this fascinating brain structure is involved in the ways that our likes and dislikes are formed, how our emotions affect our actions and memories, and how we socially interact with others. The coming decade promises to generate phenomenal further progress in specifying how associative learning and reward processing depend on representations and computations within amygdalar subregions as hr-fMRI of the amygdala becomes increasingly implemented. The future is thus incredibly bright for research on the functions of the subregions of this captivating brain structure.

REFERENCES

- Adolphs, R. (2008). Fear, faces, and the human amygdala. *Current opinion in neurobiology* 18, 166-172.
- Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences* 1191, 42-61.
- Alheid, G.F., and Heimer, L. (1988). New perspectives in basal forebrain organization of special relevance for neuropsychiatric disorders: the striatopallidal, amygdaloid, and corticopetal components of substantia innominata. *Neuroscience* 27, 1-39.
- Amano, T., Unal, C.T., and Pare, D. (2010). Synaptic correlates of fear extinction in the amygdala. *Nature neuroscience* 13, 489-494.
- Amaral, D.G., and Bassett, J.L. (1989). Cholinergic innervation of the monkey amygdala: an immunohistochemical analysis with antisera to choline acetyltransferase. *The Journal of comparative neurology* 281, 337-361.
- Amaral, D.G., Price J.L., Pitkanen, A., and Carmichael, S.T. (1992). Anatomical organization of the primate amygdaloid complex. In *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*, A. J.P., ed. (Wiley-Liss, New York), pp. 1-66.
- Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N.J., Habel, U., Schneider, F., and Zilles, K. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat Embryol (Berl)* 210, 343-352.
- Anderton, B.H. (1997). Changes in the ageing brain in health and disease. *Philos Trans R Soc Lond B Biol Sci* 352, 1781-1792.
- Angenstein, F., Kammerer, E., and Scheich, H. (2009). The BOLD response in the rat hippocampus depends rather on local processing of signals than on the input or output

activity. A combined functional MRI and electrophysiological study. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29, 2428-2439.

Bach, D.R., Behrens, T.E., Garrido, L., Weiskopf, N., and Dolan, R.J. (2011). Deep and superficial amygdala nuclei projections revealed in vivo by probabilistic tractography. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31, 618-623.

Bakker, A., Kirwan, C.B., Miller, M., and Stark, C.E. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science* 319, 1640-1642.

Ball, T., Rahm, B., Eickhoff, S.B., Schulze-Bonhage, A., Speck, O., and Mutschler, I. (2007). Response properties of human amygdala subregions: evidence based on functional MRI combined with probabilistic anatomical maps. *PloS one* 2, e307.

Balleine, B.W., and Killcross, S. (2006). Parallel incentive processing: an integrated view of amygdala function. *Trends in neurosciences* 29, 272-279.

Barger, N., Stefanacci, L., and Semendeferi, K. (2007). A comparative volumetric analysis of the amygdaloid complex and basolateral division in the human and ape brain. *American journal of physical anthropology* 134, 392-403.

Barton, R.A., and Aggleton, J.P. (2000). Primate evolution and the amygdala. In *The amygdala: a functional analysis* (Oxford University Press), pp. 479-508.

Baxter, M.G., and Murray, E.A. (2002). The amygdala and reward. *Nature reviews. Neuroscience* 3, 563-573.

Bechara, A., Damasio, H., and Damasio, A.R. (2003). Role of the amygdala in decision-making. *Annals of the New York Academy of Sciences* 985, 356-369.

Belova, M.A., Paton, J.J., Morrison, S.E., and Salzman, C.D. (2007). Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron* 55, 970-984.

- Bernard, J.F., Huang, G.F., and Besson, J.M. (1990). Effect of noxious somesthetic stimulation on the activity of neurons of the nucleus centralis of the amygdala. *Brain Res* 523, 347-350.
- Birbaumer, N., Grodd, W., Diedrich, O., Klose, U., Erb, M., Lotze, M., Schneider, F., Weiss, U., and Flor, H. (1998). fMRI reveals amygdala activation to human faces in social phobics. *Neuroreport* 9, 1223-1226.
- Bitsios, P., Szabadi, E., and Bradshaw, C.M. (2004). The fear-inhibited light reflex: importance of the anticipation of an aversive event. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology* 52, 87-95.
- Blair, H.T., Schafe, G.E., Bauer, E.P., Rodrigues, S.M., and LeDoux, J.E. (2001). Synaptic plasticity in the lateral amygdala: a cellular hypothesis of fear conditioning. *Learn Mem* 8, 229-242.
- Blundell, P., Hall, G., and Killcross, S. (2001). Lesions of the basolateral amygdala disrupt selective aspects of reinforcer representation in rats. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 21, 9018-9026.
- Boll, S., Gamer, M., Kalisch, R., and Buchel, C. (2011). Processing of facial expressions and their significance for the observer in subregions of the human amygdala. *NeuroImage* 56, 299-306.
- Bracht, T., Tuscher, O., Schnell, S., Kreher, B., Rusch, N., Glauche, V., Lieb, K., Ebert, D., Il'yasov, K.A., Hennig, J., *et al.* (2009). Extraction of prefronto-amygdalar pathways by combining probability maps. *Psychiatry research* 174, 217-222.
- Bradley, M.M. (2009). Natural selective attention: orienting and emotion. *Psychophysiology* 46, 1-11.

- Bray, S., and O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *Journal of neurophysiology* 97, 3036-3045.
- Bray, S., Rangel, A., Shimojo, S., Balleine, B., and O'Doherty, J.P. (2008). The neural mechanisms underlying the influence of pavlovian cues on human decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28, 5861-5866.
- Buchel, C., and Dolan, R.J. (2000). Classical fear conditioning in functional neuroimaging. *Current opinion in neurobiology* 10, 219-223.
- Buchel, C., and Friston, K. (2000). Assessing interactions among neuronal systems using functional neuroimaging. *Neural networks : the official journal of the International Neural Network Society* 13, 871-882.
- Buchel, C., Morris, J., Dolan, R.J., and Friston, K.J. (1998). Brain systems mediating aversive conditioning: an event-related fMRI study. *Neuron* 20, 947-957.
- Buxton, R.B. (2002). *Introduction to Functional Magnetic Resonance Imaging* (Cambridge, UK: Cambridge University Press).
- Cabeza, R., McIntosh, A.R., Tulving, E., Nyberg, L., and Grady, C.L. (1997). Age-related differences in effective neural connectivity during encoding and recall. *Neuroreport* 8, 3479-3483.
- Cador, M., Robbins, T.W., and Everitt, B.J. (1989). Involvement of the amygdala in stimulus-reward associations: interaction with the ventral striatum. *Neuroscience* 30, 77-86.
- Caesar, K., Gold, L., and Lauritzen, M. (2003). Context sensitivity of activity-dependent increases in cerebral blood flow. *Proceedings of the National Academy of Sciences of the United States of America* 100, 4239-4244.

- Cardinal, R.N., Parkinson, J.A., Hall, J., and Everitt, B.J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and biobehavioral reviews* 26, 321-352.
- Carr, V.A., Rissman, J., and Wagner, A.D. (2010). Imaging the human medial temporal lobe with high-resolution fMRI. *Neuron* 65, 298-308.
- Cheng, K., Waggoner, R.A., and Tanaka, K. (2001). Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron* 32, 359-374.
- Cohen, A., Sanborn, A., and Shiffrin, R. (2008a). Model evaluation using grouped or individual data. *Psychonomic bulletin & review* 15, 692.
- Cohen, M.X., Elger, C.E., and Weber, B. (2008b). Amygdala tractography predicts functional connectivity and learning during feedback-guided decision-making. *NeuroImage* 39, 1396-1407.
- Corbit, L.H., and Balleine, B.W. (2005). Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25, 962-970.
- Corbit, L.H., and Balleine, B.W. (2011). The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31, 11786-11794.
- Corbit, L.H., Muir, J.L., and Balleine, B.W. (2001). The role of the nucleus accumbens in instrumental conditioning: Evidence of a functional dissociation between accumbens core and shell. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 21, 3251-3260.

- Cordes, D., Turski, P.A., and Sorenson, J.A. (2000). Compensation of susceptibility-induced signal loss in echo-planar imaging for functional applications. *Magnetic resonance imaging* 18, 1055-1068.
- Corrado, G., and Doya, K. (2007). Understanding neural coding through the model-based analysis of decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27, 8178-8180.
- Courville, A.C., Daw, N.D., and Touretzky, D.S. (2006). Bayesian theories of conditioning in a changing world. *Trends in cognitive sciences* 10, 294-300.
- D'Ardenne, K., McClure, S.M., Nystrom, L.E., and Cohen, J.D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* 319, 1264-1267.
- Damasio (1994). Descartes' error.
- Davis, F.C., Johnstone, T., Mazzulla, E.C., Oler, J.A., and Whalen, P.J. (2010). Regional response differences across the human amygdaloid complex during social conditioning. *Cereb Cortex* 20, 612-621.
- Davis, M., and Shi, C. (2000). The amygdala. *Curr Biol* 10, R131.
- Davis, M., and Whalen, P.J. (2001). The amygdala: vigilance and emotion. *Molecular psychiatry* 6, 13-34.
- Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience* 8, 1704-1711.
- Dayan, P., and Daw, N.D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, affective & behavioral neuroscience* 8, 429-453.
- Dayan, P., Kakade, S., and Montague, P.R. (2000). Learning and selective attention. *Nature neuroscience* 3 *Suppl*, 1218-1223.

- Delgado, M.R., Li, J., Schiller, D., and Phelps, E.A. (2008). The role of the striatum in aversive learning and aversive prediction errors. *Philos Trans R Soc Lond B Biol Sci* 363, 3787-3800.
- Delgado, M.R., Olsson, A., and Phelps, E.A. (2006). Extending animal models of fear conditioning to humans. *Biological psychology* 73, 39-48.
- Detre, J.A., Sirven, J.I., Alsop, D.C., O'Connor, M.J., and French, J.A. (1995). Localization of subclinical ictal activity by functional magnetic resonance imaging: correlation with invasive monitoring. *Annals of neurology* 38, 618-624.
- Di Russo, F., Martinez, A., and Hillyard, S.A. (2003). Source analysis of event-related cortical activity during visuo-spatial attention. *Cereb Cortex* 13, 486-499.
- Dickinson, A., and Balleine, B.W. (2002). The role of learning in the operation of motivational systems. In *Stevens' handbook of experimental psychology*, G.R. Pashler H., ed. (New York: Wiley), pp. 497-533.
- Dolcos, F., LaBar, K.S., and Cabeza, R. (2004). Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. *Neuron* 42, 855-863.
- Donaldson, D.I. (2004). Parsing brain activity with fMRI and mixed designs: what kind of a state is neuroimaging in? *Trends in neurosciences* 27, 442-444.
- Doya, K., Samejima, K., Katagiri, K., and Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural computation* 14, 1347-1369.
- Dupont, S., Croize, A.C., Semah, F., Hasboun, D., Samson, Y., Clemenceau, S., and Baulac, M. (2001). Is amygdalohippocampectomy really selective in medial temporal lobe epilepsy? A study using positron emission tomography with (18)fluorodeoxyglucose. *Epilepsia* 42, 731-740.
- Duvernoy (2005). *The human hippocampus: Functional anatomy, vascularization and serial sections with MRI*, 3rd edition edn (New York: Springer).

Eapen, M., Zald, D.H., Gatenby, J.C., Ding, Z., and Gore, J.C. (2011). Using high-resolution MR imaging at 7T to evaluate the anatomy of the midbrain dopaminergic system. *AJNR. American journal of neuroradiology* 32, 688-694.

Ekstrom, A.D., Bazih, A.J., Suthana, N.A., Al-Hakim, R., Ogura, K., Zeineh, M., Burggren, A.C., and Bookheimer, S.Y. (2009). Advances in high-resolution imaging and computational unfolding of the human hippocampus. *NeuroImage* 47, 42-49.

Elliott, R., Newman, J.L., Longe, O.A., and William Deakin, J.F. (2004). Instrumental responding for rewards is associated with enhanced neuronal response in subcortical reward systems. *NeuroImage* 21, 984-990.

Erev, I., and Roth, A. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review* 88, 848-881.

Erk, S., Kiefer, M., Grothe, J., Wunderlich, A.P., Spitzer, M., and Walter, H. (2003). Emotional context modulates subsequent memory effect. *NeuroImage* 18, 439-447.

Everitt, B.J., Cador, M., and Robbins, T.W. (1989). Interactions between the amygdala and ventral striatum in stimulus-reward associations: studies using a second-order schedule of sexual reinforcement. *Neuroscience* 30, 63-75.

Everitt, B.J., Cardinal, R.N., Hall, J., Parkinson, J.A., and Robbins, T.W. (2000). Differential involvement of amygdala subsystems in appetitive conditioning and drug addiction. In *The Amygdala: A Functional Analysis*, J.P. Aggleton, ed. (Oxford University Press), p. p.381.

Everitt, B.J., Parkinson, J.A., Olmstead, M.C., Arroyo, M., Robledo, P., and Robbins, T.W. (1999). Associative processes in addiction and reward. The role of amygdala-ventral striatal subsystems. *Annals of the New York Academy of Sciences* 877, 412-438.

- Fanselow, M.S., and LeDoux, J.E. (1999). Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron* 23, 229-232.
- Farzaneh, F., Riederer, S.J., and Pelc, N.J. (1990). Analysis of T2 limitations and off-resonance effects on spatial resolution and artifacts in echo-planar imaging. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 14, 123-139.
- Freese, J.L., and Amaral, D.G. (2009). Neuroanatomy of the primate amygdala. In *The human amygdala*, P.E.A. Whalen P.J., ed. (The Guilford Press New York, London), pp. 3-42.
- Fried, I., Cameron, K.A., Yashar, S., Fong, R., and Morrow, J.W. (2002). Inhibitory and excitatory responses of single neurons in the human medial temporal lobe during recognition of faces and objects. *Cereb Cortex* 12, 575-584.
- Fudenberg, D., and Levine, D. (1998). Learning in games. *European economic review* 42, 631-639.
- Fudge, J.L., Kunishio, K., Walsh, P., Richard, C., and Haber, S.N. (2002). Amygdaloid projections to ventromedial striatal subterritories in the primate. *Neuroscience* 110, 257-275.
- Gallagher, M., Graham, P.W., and Holland, P.C. (1990). The amygdala central nucleus and appetitive Pavlovian conditioning: lesions impair one class of conditioned behavior. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 10, 1906-1911.
- Gallagher, M., and Holland, P.C. (1994). The amygdala complex: multiple roles in associative learning and attention. *Proceedings of the National Academy of Sciences of the United States of America* 91, 11771-11776.

- Gallagher, M., and Schoenbaum, G. (1999). Functions of the amygdala and related forebrain areas in attention and cognition. *Annals of the New York Academy of Sciences* 877, 397-411.
- Gamer, M., Zurowski, B., and Buchel, C. (2010). Different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. *Proceedings of the National Academy of Sciences of the United States of America* 107, 9400-9405.
- Garner, D.M., Olmsted, M.P., Bohr, Y., and Garfinkel, P.E. (1982). The eating attitudes test: psychometric features and clinical correlates. *Psychological medicine* 12, 871-878.
- George, J.S., Aine, C.J., Mosher, J.C., Schmidt, D.M., Ranken, D.M., Schlitt, H.A., Wood, C.C., Lewine, J.D., Sanders, J.A., and Belliveau, J.W. (1995). Mapping function in the human brain with magnetoencephalography, anatomical magnetic resonance imaging, and functional magnetic resonance imaging. *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society* 12, 406-431.
- Glascher, J., Daw, N., Dayan, P., and O'Doherty, J.P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66, 585-595.
- Glascher, J., Hampton, A.N., and O'Doherty, J.P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb Cortex* 19, 483-495.
- Glover, G.H., Li, T.Q., and Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 44, 162-167.
- Gluck, M.A., and Bower, G.H. (1988). From conditioning to category learning: an adaptive network model. *J Exp Psychol Gen* 117, 227-247.

- Gottfried, J.A., O'Doherty, J., and Dolan, R.J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 22, 10829-10837.
- Gottfried, J.A., O'Doherty, J., and Dolan, R.J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* 301, 1104-1107.
- Goulet, S., Dore, F.Y., and Murray, E.A. (1998). Aspiration lesions of the amygdala disrupt the rhinal corticothalamic projection system in rhesus monkeys. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale* 119, 131-140.
- Grill-Spector, K., Sayres, R., and Ress, D. (2006). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nature neuroscience* 9, 1177-1185.
- Groenewegen, H.J., Berendse, H.W., Wolters, J.G., and Lohman, A.H. (1990). The anatomical relationship of the prefrontal cortex with the striatopallidal system, the thalamus and the amygdala: evidence for a parallel organization. *Progress in brain research* 85, 95-116; discussion 116-118.
- Haber, S.N., and Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35, 4-26.
- Hall, J., Parkinson, J.A., Connor, T.M., Dickinson, A., and Everitt, B.J. (2001). Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behaviour. *The European journal of neuroscience* 13, 1984-1992.
- Hamann, S.B., Ely, T.D., Hoffman, J.M., and Kilts, C.D. (2002). Ecstasy and agony: activation of the human amygdala in positive and negative emotion. *Psychological science* 13, 135-141.

- Hampton, A.N., Adolphs, R., Tyszka, M.J., and O'Doherty, J.P. (2007). Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. *Neuron* 55, 545-555.
- Hampton, A.N., Bossaerts, P., and O'Doherty, J.P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26, 8360-8367.
- Hardin, M.G., Pine, D.S., and Ernst, M. (2009). The influence of context valence in the neural coding of monetary outcomes. *NeuroImage* 48, 249-257.
- Hatfield, T., Han, J.S., Conley, M., Gallagher, M., and Holland, P. (1996). Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 16, 5256-5265.
- Heimer, L. (2003). The legacy of the silver methods and the new anatomy of the basal forebrain: implications for neuropsychiatry and drug abuse. *Scandinavian journal of psychology* 44, 189-201.
- Heimer, L., Harlan, R.E., Alheid, G.F., Garcia, M.M., and de Olmos, J. (1997). Substantia innominata: a notion which impedes clinical-anatomical correlations in neuropsychiatric disorders. *Neuroscience* 76, 957-1006.
- Heimer, L., and Van Hoesen, G.W. (2006). The limbic lobe and its output channels: implications for emotional functions and adaptive behavior. *Neuroscience and biobehavioral reviews* 30, 126-147.
- Holland, P.C. (1985). The nature of conditioned inhibition in serial and simultaneous feature negative discriminations. In *Information Processing in Animals: Conditioned Inhibition*, R.R. Miller, and N.E. Spear, eds. (Hillsdale, NJ: Erlbaum), pp. 267-297.

- Holland, P.C. (2004). Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *Journal of experimental psychology. Animal behavior processes* 30, 104-117.
- Holland, P.C., and Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. *Trends in cognitive sciences* 3, 65-73.
- Holland, P.C., and Gallagher, M. (2003). Double dissociation of the effects of lesions of basolateral and central amygdala on conditioned stimulus-potentiated feeding and Pavlovian-instrumental transfer. *The European journal of neuroscience* 17, 1680-1694.
- Holland, P.C., and Gallagher, M. (2004). Amygdala-frontal interactions and reward expectancy. *Current opinion in neurobiology* 14, 148-155.
- Holland, P.C., Petrovich, G.D., and Gallagher, M. (2002). The effects of amygdala lesions on conditioned stimulus-potentiated eating in rats. *Physiol Behav* 76, 117-129.
- Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C.F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310, 1680-1683.
- Iidaka, T., Omori, M., Murata, T., Kosaka, H., Yonekura, Y., Okada, T., and Sadato, N. (2001). Neural interaction of the amygdala with the prefrontal and temporal cortices in the processing of facial expressions as revealed by fMRI. *Journal of cognitive neuroscience* 13, 1035-1047.
- Insausti, A. (2004). Hippocampal formation. In *Human nervous system*, J. Paxinos, ed. (San Diego: Elsevier Academic Press), pp. 871-913.
- Insausti, R., Juottonen, K., Soininen, H., Insausti, A.M., Partanen, K., Vainio, P., Laakso, M.P., and Pitkanen, A. (1998). MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *AJNR. American journal of neuroradiology* 19, 659-671.

- Irwin, W., Davidson, R.J., Lowe, M.J., Mock, B.J., Sorenson, J.A., and Turski, P.A. (1996). Human amygdala activation detected with echo-planar functional magnetic resonance imaging. *Neuroreport* 7, 1765-1769.
- Johansen, J.P., Cain, C.K., Ostroff, L.E., and LeDoux, J.E. (2011). Molecular mechanisms of fear learning and memory. *Cell* 147, 509-524.
- Kalin, N.H. (2004). Studying non-human primates: a gateway to understanding anxiety disorders. *Psychopharmacology bulletin* 38 *Suppl* 1, 8-13.
- Kamin, L.J. (1969). Selective association and conditioning. In *Fundamental Issues in Associative Learning*, N.J. Mackintosh, and W.K. Honig, eds. (Halifax, Nova Scotia: Dalhousie Univ. Press), pp. 42-64.
- Kelley, A.E., Domesick, V.B., and Nauta, W.J. (1982). The amygdalostratial projection in the rat--an anatomical study by anterograde and retrograde tracing methods. *Neuroscience* 7, 615-630.
- Kensinger, E.A., and Schacter, D.L. (2006). Amygdala activity is associated with the successful encoding of item, but not source, information for positive and negative stimuli. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26, 2564-2570.
- Killcross, S., Robbins, T.W., and Everitt, B.J. (1997). Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. *Nature* 388, 377-380.
- Kim, H., Shimojo, S., and O'Doherty, J.P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol* 4, e233.
- Kim, H., Somerville, L.H., Johnstone, T., Alexander, A.L., and Whalen, P.J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *Neuroreport* 14, 2317-2322.

- Kim, H., Somerville, L.H., Johnstone, T., Polis, S., Alexander, A.L., Shin, L.M., and Whalen, P.J. (2004). Contextual modulation of amygdala responsivity to surprised faces. *Journal of cognitive neuroscience* 16, 1730-1745.
- Kinahan, P.E., and Noll, D.C. (1999). A direct comparison between whole-brain PET and BOLD fMRI measurements of single-subject activation response. *NeuroImage* 9, 430-438.
- Kirwan, C.B., Jones, C.K., Miller, M.I., and Stark, C.E. (2007). High-resolution fMRI investigation of the medial temporal lobe. *Human brain mapping* 28, 959-966.
- Koenigs, M., and Grafman, J. (2009). Posttraumatic stress disorder: the role of medial prefrontal cortex and amygdala. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* 15, 540-548.
- Konorski, J. (1967). *Integrative Activity of the Brain* (University of Chicago Press).
- Krasnow, B., Tamm, L., Greicius, M.D., Yang, T.T., Glover, G.H., Reiss, A.L., and Menon, V. (2003). Comparison of fMRI activation at 3 and 1.5 T during perceptual, cognitive, and affective processing. *NeuroImage* 18, 813-826.
- Krettek, J.E., and Price, J.L. (1977). Projections from the amygdaloid complex to the cerebral cortex and thalamus in the rat and cat. *The Journal of comparative neurology* 172, 687-722.
- Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., and Baker, C.I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature neuroscience* 12, 535-540.
- LaBar, K.S., Gatenby, J.C., Gore, J.C., LeDoux, J.E., and Phelps, E.A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron* 20, 937-945.
- LaBar, K.S., Gitelman, D.R., Mesulam, M.M., and Parrish, T.B. (2001). Impact of signal-to-noise on functional MRI of the human amygdala. *Neuroreport* 12, 3461-3464.

- Lai, S., Hopkins, A.L., Haacke, E.M., Li, D., Wasserman, B.A., Buckley, P., Friedman, L., Meltzer, H., Hedera, P., and Friedland, R. (1993). Identification of vascular structures as a major source of signal contrast in high resolution 2D and 3D functional activation imaging of the motor cortex at 1.5T: preliminary results. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 30, 387-392.
- LeDoux, J. (1998). Fear and the brain: where have we been, and where are we going? *Biological psychiatry* 44, 1229-1238.
- LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cell Mol Neurobiol* 23, 727-738.
- LeDoux, J. (2007). The amygdala. *Curr Biol* 17, R868-874.
- LeDoux, J.E. (2000a). The amygdala and emotion: A view through fear. In *The Amygdala: A functional analysis*, A. J.P., ed. (Oxford Univeristy Press), pp. 289-310.
- LeDoux, J.E. (2000b). Emotion circuits in the brain. *Annual review of neuroscience* 23, 155-184.
- LeDoux, J.E., Cicchetti, P., Xagoraris, A., and Romanski, L.M. (1990). The lateral amygdaloid nucleus: sensory interface of the amygdala in fear conditioning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 10, 1062-1069.
- Li, J., Schiller, D., Schoenbaum, G., Phelps, E.A., and Daw, N.D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature neuroscience* 14, 1250-1252.
- Libby, W.L., Jr., Lacey, B.C., and Lacey, J.I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology* 10, 270-294.
- Logothetis, N.K. (2008). What we can do and what we cannot do with fMRI. *Nature* 453, 869-878.

- Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150-157.
- Logothetis, N.K., and Wandell, B.A. (2004). Interpreting the BOLD signal. *Annual review of physiology* 66, 735-769.
- Lovibond, P.F. (1983). Facilitation of instrumental behavior by a Pavlovian appetitive conditioned stimulus. *Journal of experimental psychology. Animal behavior processes* 9, 225-247.
- Lubow, R.E., and Moore, A.U. (1959). Latent inhibition: the effect of nonreinforced pre-exposure to the conditional stimulus. *Journal of comparative and physiological psychology* 52, 415-419.
- Luce, R.D. (1959). On the possible psychophysical laws. *Psychological review* 66, 81-95.
- Machii, K., Cohen, D., Ramos-Estebanez, C., and Pascual-Leone, A. (2006). Safety of rTMS to non-motor cortical areas in healthy participants and patients. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* 117, 455-471.
- Mackiewicz, K.L., Sarinopoulos, I., Cleven, K.L., and Nitschke, J.B. (2006). The effect of anticipation and the specificity of sex differences for amygdala and hippocampus function in emotional memory. *Proceedings of the National Academy of Sciences of the United States of America* 103, 14200-14205.
- Mackintosh, N.J. (1975). A theory of attention: variations on the associability of stimuli with reinforcement *Psychol. Rev.* , 276-298.
- Magistretti, P.J., and Pellerin, L. (1999). Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. *Philos Trans R Soc Lond B Biol Sci* 354, 1155-1163.

- Mai, J.K., Paxinos, G., and Voss, T. (2008). Atlas of the human brain, 3rd Edition edn (Elsevier).
- Maren, S. (2005). Building and burying fear memories in the brain. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* 11, 89-99.
- Markowitsch, H.J., Calabrese, P., Wurker, M., Durwen, H.F., Kessler, J., Babinsky, R., Brechtelsbauer, D., Heuser, L., and Gehlen, W. (1994). The amygdala's contribution to memory--a study on two patients with Urbach-Wiethe disease. *Neuroreport* 5, 1349-1352.
- Martinez, A., Anllo-Vento, L., Sereno, M.I., Frank, L.R., Buxton, R.B., Dubowitz, D.J., Wong, E.C., Hinrichs, H., Heinze, H.J., and Hillyard, S.A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature neuroscience* 2, 364-369.
- McClure, S.M., Berns, G.S., and Montague, P.R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38, 339-346.
- McDonald, A.J. (1991a). Organization of amygdaloid projections to the prefrontal cortex and associated striatum in the rat. *Neuroscience* 44, 1-14.
- McDonald, A.J. (1991b). Topographical organization of amygdaloid projections to the caudatoputamen, nucleus accumbens, and related striatal-like areas of the rat brain. *Neuroscience* 44, 15-33.
- McDonald, A.J. (1998). Cortical pathways to the mammalian amygdala. *Progress in neurobiology* 55, 257-332.
- Milad, M.R., Pitman, R.K., Ellis, C.B., Gold, A.L., Shin, L.M., Lasko, N.B., Zeidan, M.A., Handwerker, K., Orr, S.P., and Rauch, S.L. (2009). Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biological psychiatry* 66, 1075-1082.

- Miller, M.I., Beg, M.F., Ceritoglu, C., and Stark, C. (2005). Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. *Proceedings of the National Academy of Sciences of the United States of America* 102, 9685-9690.
- Mobbs, D., Petrovic, P., Marchant, J.L., Hassabis, D., Weiskopf, N., Seymour, B., Dolan, R.J., and Frith, C.D. (2007). When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science* 317, 1079-1083.
- Morris, J.S., Buchel, C., and Dolan, R.J. (2001a). Parallel neural responses in amygdala subregions and sensory cortex during implicit fear conditioning. *NeuroImage* 13, 1044-1052.
- Morris, J.S., deBonis, M., and Dolan, R.J. (2002). Human amygdala responses to fearful eyes. *NeuroImage* 17, 214-222.
- Morris, J.S., DeGelder, B., Weiskrantz, L., and Dolan, R.J. (2001b). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain : a journal of neurology* 124, 1241-1252.
- Morris, J.S., and Dolan, R.J. (2004). Dissociable amygdala and orbitofrontal responses during reversal fear conditioning. *NeuroImage* 22, 372-380.
- Morris, J.S., Friston, K.J., Buchel, C., Frith, C.D., Young, A.W., Calder, A.J., and Dolan, R.J. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain : a journal of neurology* 121 (Pt 1), 47-57.
- Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., and Malach, R. (2005). Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* 309, 951-954.
- Mulert, C., Jager, L., Schmitt, R., Bussfeld, P., Pogarell, O., Moller, H.J., Juckel, G., and Hegerl, U. (2004). Integration of fMRI and simultaneous EEG: towards a

comprehensive understanding of localization and time-course of brain activity in target detection. *NeuroImage* 22, 83-94.

Murray, E.A. (2007). The amygdala, reward and emotion. *Trends in cognitive sciences* 11, 489-497.

Myung, I., and Busemeyer, J. (1992). Measurement-free tests of a general state-space model of prototype learning. *Journal of Mathematical Psychology* 36, 32-67.

Nelder, J., and Mead, R. (1965). A simplex method for function minimization. *The computer journal* 7, 308.

Newton, F.H., Rosenberg, R.N., Lampert, P.W., and O'Brien, J.S. (1971). Neurologic involvement in Urbach-Wiethe's disease (lipoid proteinosis). A clinical, ultrastructural, and chemical study. *Neurology* 21, 1205-1213.

Nicotra, A., Critchley, H.D., Mathias, C.J., and Dolan, R.J. (2006). Emotional and autonomic consequences of spinal cord injury explored using functional brain imaging. *Brain : a journal of neurology* 129, 718-728.

O'Brien, C.P., Childress, A.R., Ehrman, R., and Robbins, S.J. (1998). Conditioning factors in drug abuse: can they explain compulsion? *J Psychopharmacol* 12, 15-22.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452-454.

O'Doherty, J., Rolls, E.T., Francis, S., Bowtell, R., and McGlone, F. (2001). Representation of pleasant and aversive taste in the human brain. *Journal of neurophysiology* 85, 1315-1321.

O'Doherty, J.P., Hampton, A., and Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences* 1104, 35-53.

- Parkinson, J.A., Robbins, T.W., and Everitt, B.J. (2000). Dissociable roles of the central and basolateral amygdala in appetitive emotional learning. *The European journal of neuroscience* 12, 405-413.
- Paton, J.J., Belova, M.A., Morrison, S.E., and Salzman, C.D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* 439, 865-870.
- Pearce, J.M., and Bouton, M.E. (2001). Theories of associative learning in animals. *Annual review of psychology* 52, 111-139.
- Pearce, J.M., and Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological review* 87, 532-552.
- Petrovich, G.D., Ross, C.A., Mody, P., Holland, P.C., and Gallagher, M. (2009). Central, but not basolateral, amygdala is critical for control of feeding by aversive learned cues. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29, 15205-15212.
- Phelps, E.A. (2006). Emotion and cognition: insights from studies of the human amygdala. *Annual review of psychology* 57, 27-53.
- Phelps, E.A., and LeDoux, J.E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* 48, 175-187.
- Phillips, R.G., and LeDoux, J.E. (1992). Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behavioral neuroscience* 106, 274-285.
- Pickens, C.L., Saddoris, M.P., Setlow, B., Gallagher, M., Holland, P.C., and Schoenbaum, G. (2003). Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23, 11078-11084.

- Pinker (1998). *How the mind works*. (Penguin: London).
- Pitkänen, A. (2000). Connectivity of the rat amygdaloid complex. In *The amygdala: A functional analysis*, Aggleton, J.P., ed. (Oxford University Press).
- Pitkänen, A., and Amaral, D.G. (1998). Organization of the intrinsic connections of the monkey amygdaloid complex: projections originating in the lateral nucleus. *The Journal of comparative neurology* 398, 431-458.
- Pitkänen, A., and Kempainen, S. (2002). Comparison of the distribution of calcium-binding proteins and intrinsic connectivity in the lateral nucleus of the rat, monkey, and human amygdala. *Pharmacology, biochemistry, and behavior* 71, 369-377.
- Pitkänen, A., Savander, V., and LeDoux, J.E. (1997). Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. *Trends in neurosciences* 20, 517-523.
- Posner, M.I., Petersen, S.E., Fox, P.T., and Raichle, M.E. (1988). Localization of cognitive operations in the human brain. *Science* 240, 1627-1631.
- Prevost, C., McCabe, J.A., Jessup, R.K., Bossaerts, P., and O'Doherty, J.P. (2011). Differentiable contributions of human amygdalar subregions in the computations underlying reward and avoidance learning. *The European journal of neuroscience* 34, 134-145.
- Price, J.L. (1986). Subcortical projections from the amygdaloid complex. *Advances in experimental medicine and biology* 203, 19-33.
- Price, J.L. (2003). Comparative aspects of amygdala connectivity. *Annals of the New York Academy of Sciences* 985, 50-58.
- Price, J.L., and Amaral, D.G. (1981). An autoradiographic study of the projections of the central nucleus of the monkey amygdala. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 1, 1242-1259.

- Pruessner, J.C., Kohler, S., Crane, J., Pruessner, M., Lord, C., Byrne, A., Kabani, N., Collins, D.L., and Evans, A.C. (2002). Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. *Cereb Cortex* 12, 1342-1353.
- Pruessner, J.C., Li, L.M., Serles, W., Pruessner, M., Collins, D.L., Kabani, N., Lupien, S., and Evans, A.C. (2000). Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: minimizing the discrepancies between laboratories. *Cereb Cortex* 10, 433-442.
- Puce, A., Constable, R.T., Luby, M.L., McCarthy, G., Nobre, A.C., Spencer, D.D., Gore, J.C., and Allison, T. (1995). Functional magnetic resonance imaging of sensory and motor cortex: comparison with electrophysiological localization. *Journal of neurosurgery* 83, 262-270.
- Rauch, S.L., Shin, L.M., and Phelps, E.A. (2006). Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research--past, present, and future. *Biological psychiatry* 60, 376-382.
- Rescorla, R.A. (1972). A theory of Pavlovian conditioning: Variations in effectiveness of reinforcement and nonreinforcement. . In *Classical conditioning II: Current research and Theory*, N.-Y.A. Century-Crofts, ed. (Black A.H, P.W.H.).
- Rescorla, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In *Classical conditioning: Current research and theory*, A.H.B.W.F. Proksty, ed. (New York: Appleton-Century-Crofts), pp. 64-99.
- Rodrigues, S.M., Schafe, G.E., and LeDoux, J.E. (2004). Molecular mechanisms underlying emotional learning and memory in the lateral amygdala. *Neuron* 44, 75-91.

Roesch, M.R., Calu, D.J., Esber, G.R., and Schoenbaum, G. (2010). Neural correlates of variations in event processing during learning in basolateral amygdala. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30, 2464-2471.

Rogan, M.T., and LeDoux, J.E. (1996). Emotion: systems, cells, synaptic plasticity. *Cell* 85, 469-475.

Roy, C.S., and Sherrington, C.S. (1890). On the Regulation of the Blood-supply of the Brain. *The Journal of physiology* 11, 85-158 117.

Russchen, F.T., Bakst, I., Amaral, D.G., and Price, J.L. (1985). The amygdalostriatal projections in the monkey. An anterograde tracing study. *Brain Res* 329, 241-257.

Saddoris, M.P., Gallagher, M., and Schoenbaum, G. (2005). Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. *Neuron* 46, 321-331.

Saygin, Z.M., Osher, D.E., Augustinack, J., Fischl, B., and Gabrieli, J.D. (2011). Connectivity-based segmentation of human amygdala nuclei using probabilistic tractography. *NeuroImage* 56, 1353-1361.

Schiller, D., Levy, I., Niv, Y., LeDoux, J.E., and Phelps, E.A. (2008). From fear to safety and back: reversal of fear in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28, 11517-11525.

Schoenbaum, G., Chiba, A.A., and Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature neuroscience* 1, 155-159.

Schoenbaum, G., Chiba, A.A., and Gallagher, M. (1999). Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 19, 1876-1884.

- Schoenbaum, G., Chiba, A.A., and Gallagher, M. (2000). Changes in functional connectivity in orbitofrontal cortex and basolateral amygdala during learning and reversal training. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 20, 5179-5189.
- Schoenbaum, G., and Roesch, M. (2005). Orbitofrontal cortex, associative learning, and expectancies. *Neuron* 47, 633-636.
- Schoenbaum, G., Setlow, B., Nugent, S.L., Saddoris, M.P., and Gallagher, M. (2003a). Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. *Learn Mem* 10, 129-140.
- Schoenbaum, G., Setlow, B., Saddoris, M.P., and Gallagher, M. (2003b). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* 39, 855-867.
- Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron* 36, 241-263.
- Schwarz (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.
- Seymour, B., Daw, N., Dayan, P., Singer, T., and Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27, 4826-4831.
- Seymour, B., and Dolan, R. (2008). Emotion, decision making, and the amygdala. *Neuron* 58, 662-671.
- Seymour, B., O'Doherty, J.P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., and Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature neuroscience* 8, 1234-1240.
- Shabel, S.J., and Janak, P.H. (2009). Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. *Proceedings of the National Academy of Sciences of the United States of America* 106, 15031-15036.

- Shi, C.J., and Cassell, M.D. (1997). Cortical, thalamic, and amygdaloid projections of rat temporal cortex. *The Journal of comparative neurology* 382, 153-175.
- Shi, C.J., and Cassell, M.D. (1998a). Cascade projections from somatosensory cortex to the rat basolateral amygdala via the parietal insular cortex. *The Journal of comparative neurology* 399, 469-491.
- Shi, C.J., and Cassell, M.D. (1998b). Cortical, thalamic, and amygdaloid connections of the anterior and posterior insular cortices. *The Journal of comparative neurology* 399, 440-468.
- Shi, C.J., and Cassell, M.D. (1999). Perirhinal cortex projections to the amygdaloid complex and hippocampal formation in the rat. *The Journal of comparative neurology* 406, 299-328.
- Small, S.A., Nava, A.S., Perera, G.M., Delapaz, R., and Stern, Y. (2000a). Evaluating the function of hippocampal subregions with high-resolution MRI in Alzheimer's disease and aging. *Microscopy research and technique* 51, 101-108.
- Small, S.A., Wu, E.X., Bartsch, D., Perera, G.M., Lacefield, C.O., DeLaPaz, R., Mayeux, R., Stern, Y., and Kandel, E.R. (2000b). Imaging physiologic dysfunction of individual hippocampal subregions in humans and genetically modified mice. *Neuron* 28, 653-664.
- Smeets, P.M., and Barnes-Holmes, D. (2003). Children's emergent preferences for soft drinks: stimulus-equivalence and transfer. *J Econ Psychol*, 603-618.
- Solano-Castiella, E., Anwender, A., Lohmann, G., Weiss, M., Docherty, C., Geyer, S., Reimer, E., Friederici, A.D., and Turner, R. (2010). Diffusion tensor imaging segments the human amygdala in vivo. *NeuroImage* 49, 2958-2965.
- Sorvari, H., Soininen, H., Paljarvi, L., Karkola, K., and Pitkanen, A. (1995). Distribution of parvalbumin-immunoreactive cells and fibers in the human amygdaloid complex. *The Journal of comparative neurology* 360, 185-212.

- Sorvari, H., Soininen, H., and Pitkanen, A. (1996a). Calbindin-D28K-immunoreactive cells and fibres in the human amygdaloid complex. *Neuroscience* 75, 421-443.
- Sorvari, H., Soininen, H., and Pitkanen, A. (1996b). Calretinin-immunoreactive cells and fibers in the human amygdaloid complex. *The Journal of comparative neurology* 369, 188-208.
- Stalnaker, T.A., Franz, T.M., Singh, T., and Schoenbaum, G. (2007). Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. *Neuron* 54, 51-58.
- Stephan, H., Frahm, H.D., and Baron, G. (1987). Comparison of brain structure volumes in Insectivora and primates. VII. Amygdaloid components. *Journal fur Hirnforschung* 28, 571-584.
- Sun, P., Ueno, K., Waggoner, R.A., Gardner, J.L., Tanaka, K., and Cheng, K. (2007). A temporal frequency-dependent functional architecture in human V1 revealed by high-resolution fMRI. *Nature neuroscience* 10, 1404-1406.
- Sutton, R., and Barto, A. (1998). *Reinforcement learning* (MIT Press).
- Swanson, L.W., and Petrovich, G.D. (1998). What is the amygdala? *Trends in neurosciences* 21, 323-331.
- Talairach, T. (1988). *Co-planar stereotaxic atlas of the human brain* (New York: Thieme).
- Talmi, D., Seymour, B., Dayan, P., and Dolan, R.J. (2008). Human pavlovian-instrumental transfer. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28, 360-368.
- Taylor, S.F., Liberzon, I., and Koeppe, R.A. (2000). The effect of graded aversive stimuli on limbic and visual activation. *Neuropsychologia* 38, 1415-1425.
- Thirion, J.P. (1998). Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical image analysis* 2, 243-260.

- Tranel, D., and Hyman, B.T. (1990). Neuropsychological correlates of bilateral amygdala damage. *Archives of neurology* 47, 349-355.
- Turner, B.H., and Herkenham, M. (1991). Thalamoamygdaloid projections in the rat: a test of the amygdala's role in sensory processing. *The Journal of comparative neurology* 313, 295-325.
- Tuunanen, J., Halonen, T., and Pitkanen, A. (1996). Status epilepticus causes selective regional damage and loss of GABAergic neurons in the rat amygdaloid complex. *The European journal of neuroscience* 8, 2711-2725.
- Uttal (2001). *The New Phrenology: The limits of Localising Cognitive Processes in the Brain* (Cambridge, MA: MIT Press).
- Vazquez, A.L., and Noll, D.C. (1998). Nonlinear aspects of the BOLD response in functional MRI. *NeuroImage* 7, 108-118.
- Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2007). Non-parametric diffeomorphic image registration with the demons algorithm. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* 10, 319-326.
- Visscher, K.M., Miezin, F.M., Kelly, J.E., Buckner, R.L., Donaldson, D.I., McAvoy, M.P., Bhalodia, V.M., and Petersen, S.E. (2003). Mixed blocked/event-related designs separate transient and sustained activity in fMRI. *NeuroImage* 19, 1694-1708.
- Wagner, A.R. (1981). SOP: a model of automatic memory processing in animal behaviour. In *Information Processing in Animals: Memory mechanisms*, N.E. Spear, and R.R. Miller, eds. (Hillsdale, NJ: Erlbaum), pp. 5-47.
- Whalen, P.J., Rauch, S.L., Etcoff, N.L., McInerney, S.C., Lee, M.B., and Jenike, M.A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 18, 411-418.

- Whalen, P.J., Shin, L.M., McInerney, S.C., Fischer, H., Wright, C.I., and Rauch, S.L. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion* 1, 70-83.
- Yacubian, J., Glascher, J., Schroeder, K., Sommer, T., Braus, D.F., and Buchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26, 9530-9537.
- Yamasaki, H., LaBar, K.S., and McCarthy, G. (2002). Dissociable prefrontal brain systems for attention and emotion. *Proceedings of the National Academy of Sciences of the United States of America* 99, 11447-11451.
- Yassa, M.A., and Stark, C.E. (2009). A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. *NeuroImage* 44, 319-327.
- Young, T.M., Asahina, M., Nicotra, A., and Mathias, C.J. (2006). Skin vasomotor reflex responses in two contrasting groups of autonomic failure: multiple system atrophy and pure autonomic failure. *Journal of neurology* 253, 846-850.
- Zald, D.H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain research. Brain research reviews* 41, 88-123.
- Zald, D.H., Donndelinger, M.J., and Pardo, J.V. (1998). Elucidating dynamic brain interactions with across-subjects correlational analyses of positron emission tomographic data: the functional connectivity of the amygdala and orbitofrontal cortex during olfactory tasks. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism* 18, 896-905.
- Zald, D.H., Hagen, M.C., and Pardo, J.V. (2002). Neural correlates of tasting concentrated quinine and sugar solutions. *Journal of neurophysiology* 87, 1068-1075.

Zald, D.H., and Pardo, J.V. (2002). The neural correlates of aversive auditory stimulation. *NeuroImage* 16, 746-753.

Zeineh, M.M., Engel, S.A., and Bookheimer, S.Y. (2000). Application of cortical unfolding techniques to functional MRI of the human hippocampal region. *NeuroImage* 11, 668-683.