THE CAPGRAS DELUSION

AN INTEGRATED APPROACH

Neralie Diane Wise

BA (Hon) Philosophy Macquarie University NSW Australia

This thesis is presented for the degree of Doctor of Philosophy submitted in the Philosophy Department, Faculty of Arts
Macquarie University

August 2012

Table of Contents

TABLE OF CONTENTS ..... **II**

TABLE OF FIGURES ..... **VI**

ABSTRACT ..... **VII**

STATEMENT OF CANDIDATE ..... **VIII**

ACKNOWLEDGEMENTS ..... **IX**

INTRODUCTION **1**

1. Introduction *1*

2. Thesis structure *5*

3. Part One *6*

4. Part Two *13*

PART ONE **20**

CHAPTER 1: PHENOMENOLOGY AND DELUSIONS **21**

1. Introduction *21*

SECTION ONE: KARL JASPERS 21

1. Introduction *21*

2. Concepts *22*

i) The psyche 22

ii) Understanding 24

iii) Form and content 25

iv) The issue of reality 26

3. Delusions *27*

4. Conclusion *30*

SECTION TWO: FOLLOWING JASPERS 31

1. Introduction *31*

2. Louis Sass *32*

i) Hyperreflexivity 33

ii) Body alienation 34

iii) Un-worlding 34

iv) Subjectivization 35

3. Shaun Gallagher *37*

i) Embodied and embedded cognition 38

ii) How the Body Shapes the Mind 39

iii) The multiple realities hypothesis 42

4. Conclusion *45*

CHAPTER 2: BRENDAN MAHER **47**

1. Introduction *47*

SECTION ONE: BRENDAN MAHER 47

1. Maher's Challenge to Jaspers *47*

2. Maher's Arguments *49*

i) The anomalous experience 49

ii) The question of rationality 50

iii) Evolution and cognition 51

iv) Impaired reasoning 52

3. Criticisms and Responses *52*

i) Why do most people fail to develop delusions when faced with anomalous experiences? 52

ii) How does the model account for delusions that occur in the absence of any anomalous experience? 53

iii) Why do patients develop a delusional explanation rather than a normal one? 54

iv) Why do patients, unlike normal people, fail to reject the explanation? 54

4. Summary and Conclusion *55*

SECTION TWO: TESTING MAHER'S TWO CLAIMS 55

1. Introduction *56*
2. Biological Causes of Delusion *56*
i) Stroke, tumour, trauma, dementing illness and disease 56
ii) Neuroimaging 57
3. The Patient's Thinking is Normal *58*
i) Attributional style 59
ii) Conservation and observational adequacy 60
iii) Probabilistic reasoning bias 61
4. Summary and Conclusion *64*

CHAPTER 3: TWO-FACTOR ACCOUNTS **66**
1. Introduction *66*
2. Two-factor accounts *66*
3. Max Coltheart's two deficits theory *68*
i) Factor one (deficit A) 68
ii) Factor two (deficit B) 73
iii) Nonneuropsychological delusions 74
iv) Discussion 76
4. Anne Davies and Martin Davies – parametric variations *77*
i) The first factor 78
ii) The second factor 79
iii) Discussion 81
5. Conclusion *82*

CHAPTER 4: ONE-FACTOR ACCOUNTS **84**
1. Introduction *84*
2. Phil Gerrans –Dopamine dysregulation *85*
i) Online and offline cognition 85
ii) Memory and imagination 87
iii) DA dysregulation 89
iv) Summary and conclusion 91
4. Fletcher and Colleagues' Prediction Error Theory of Delusions *94*
i) Prediction error learning 94
ii) The Bayesian brain 96
iii) Neuropsychopharmacology 98
iv) Capgras and Fregoli 104
v) Discussion 105

CHAPTER 5: TWO CONTENTIOUS ISSUES **109**
1. Introduction *109*
SECTION ONE: EXPERIENCE AND BELIEFS 109
1. Introduction *109*
2. Non-doxastic accounts *110*
i) Berrios 111
ii) Currie 111
iii) Sass 112
3. Doxastic accounts: explanationist v/s endorsement *113*
i) The explanationist account 114
ii) The endorsement account 118
4. Two Ways of Resolving the Differences *121*
i) Max Coltheart 121
ii) Langdon and Bayne 123
5. Conclusion *126*
SECTION TWO: THE CONTENT OF PERCEPTION 126
1. Introduction *126*
2. Defining Perception and its Contents *127*
3. Jerry Fodor *128*
4. Paul Churchland *130*
5. Athanassios Raftopoulos *131*

6. Shaun Gallagher *134*
7. Perception, Meaning and Concepts *138*

PART TWO **140**

INTRODUCTION **141**

CHAPTER 6: THE CAPGRAS DELUSION: **145**
1. Introduction *145*
2. The Capgras Delusion *146*
3. The Delayed-Processing Account *148*
i) Whittlesea and Williams 148
ii) The Capgras delusion 151
4. Accounting for the SCR Data *154*
5. The Tendency of the Delusion to Spread Over Time *156*
6. Delayed-processing and the Erroneous Classification *158*
7. Capgras for Inanimate Objects *159*
8. Summary and Conclusion *162*

CHAPTER 7: MEANING AND THE CAPGRAS DELUSION **164**
1. Introduction *164*
2. The phenomenology *164*
3. Objects, contexts, scenes and scenarios *166*
4. The explanationist account *169*
5. Misidentification and intersubjectivity *170*
i) Primary intersubjectivity 172
ii) Secondary Intersubjectivity 174
iii) Narrative competency 174
6. The Attribution Process *176*
7. The 'impostor' attribution *178*

CHAPTER 8: ALTERNATIVE REALITIES AND DELUSIONS **180**
1. Introduction *180*
2. Delusions and alternative realities *180*
3. Creating imaginary worlds *182*
4. Make-believe and the Capgras delusion *187*

CHAPTER 9: EXPLAINING UNUSUAL CHARACTERISTICS **192**
1. Introduction *192*
2. Passivity and aggression *192*
3. Circumscription and double-bookkeeping *200*
4. Bizarre content *204*
5. Incorrigibility *204*
6. Incomplete worlds and the unexplained *205*
7. Conclusion *207*

CHAPTER 10: THE CAPGRAS DELUSION **208**
1. Introduction *208*
2. Source monitoring *209*
3. Keeping different worlds apart *210*
4. Hypnosis and delusions *214*
i) Imagination and social context 216
ii) Dissociation 217
5. Conflict monitoring, executive control and neuroimaging *219*
6. Decoupling: a useful mechanism *221*
7. Capgras delusion: the second factor *223*
8. In reply to Hohwy and Rajan *226*
9. Ventromedial frontal lesions *228*

10. Testing the hypothesis *229*

CHAPTER 11: FURTHER APPLICATIONS **231**

1. Introduction *231*

2. The alternative reality of the imagination *231*

i) Mirrored-self misidentification without mirror agnosia 231

ii) Thought insertion 234

iii) Thought control 236

3. The alternative realities of cognitive levels *237*

i) Mirrored-self misidentification with mirror agnosia 238

ii) Somatoparaphrenia 239

4. Schizophrenia *244*

SUMMARY AND CONCLUSION **247**

BIBLIOGRAPHY **251**

Table of Figures

Figure 1: The Ellis face processing model shows how a lesion at A could result in prosopagnosia, and a lesion at B result in the Capgras delusion. 148

Figure 2: A and B represent the opposing ends of a continuum. C and D represent the sites of impairment with C leading to prosopagnosia and D to the feeling of familiarity described by Whittlesea and Williams. The discrepancy is between the classification of the face as 'unfamiliar' and the accessing of the PIN. 152

Figure 3: (A) Structural encoding slower than normally would be expected for a specific face causes the face to be falsely classified as 'unfamiliar'. (B) Deterioration of the condition (increased reduction in speed) causes less familiar faces to be falsely classified. (C) Extreme condition artificially produced in the laboratory, in which all, or most, familiar faces are falsely classified. 157

Figure 4: A. An 'unfamiliar classification is not followed by recognition so no feedback loop develops. B. Processes that repeatedly occur together enable feedback loops to develop. 158

Figure 5: An attempt to insert a concept from a fictional world into the real world causes a cross-world conflict. The conflict monitor alerts executive control and stimulates a phenomenal response, which engages conscious awareness. As a result the conflict is resolved (error corrected) 223

Figure 6: The vicious cycle created by the delayed face encoding in the Capgras delusion 224

Figure 7: To avoid the vicious cycle, involuntary decoupling of the conflict monitor and executive control occurs to enable the fiction to be experienced as real, thus cohering with the experience of the everyday world as real 225

Abstract

Delusions are studied in two different philosophical traditions: the continental or phenomenological tradition and the Anglo-American or analytic tradition. This thesis argues that the most plausible account of monothematic delusions ultimately incorporates valuable insights from both. It offers a new two-factor account of the Capgras delusion in which the characterisation of each factor reflects insights from the two traditions. I argue that the first factor is a delay in face *encoding,* which causes the face to be classified 'unknown', thus creating a conflict when the person's identity is accessed. I argue that this conflict is resolved by an application of the 'impostor' concept, which is an interpretation of the face *in context*, and further, that this concept is primarily developed through acquaintance with fiction. Applying a concept from fiction to reality creates cross-world conflict that cannot be resolved in the normal way. Consequently, the system that monitors conflicts is decoupled from executive control to allow the impostor attribution to pass unchallenged. The decoupling constitutes the second factor. I argue that the decoupling, which is a normal process that allows engagement with imaginative worlds, is brought to bear on an abnormal situation. This two-factor account is supported by neuroimaging data from investigations into conflict monitoring and executive control, and investigation of the hypnotic condition. In conclusion, I will suggest how the model presented could be applied to several other monothematic delusions.

Statement of Candidate

I certify that the work in this thesis entitled "The Capgras Delusion: An Integrated Account" has not been previously submitted for a degree nor has it been submitted as part of requirements for a degree to any other University or institution other than Macquarie University.

I also certify that this thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis.

Neralie Diane Wise (Student ID 40223132)
August 2012

Acknowledgements

I would like to express my appreciation to the following supervisors for their support and guidance during my candidature.

Prof. Peter Menzies: principal supervisor from the second half of the first year until thesis submission. I would like to thank Prof. Peter Menzies for his excellent supervision, for asking me the hard questions, and for persuading me to pursue challenges I would have otherwise avoided. I would also like to thank him for his personal support during my treatment for cancer.

Dr. Mitch Parsell: principal supervisor for the first six months, and associate supervisor for the remainder of my candidature. I would like to thank Dr. Mitch Parsell for being there for my entire candidature, for always being supportive, encouraging and generous with his time. I would also like to thank him for being a friend when times were challenging.

Dr. Robert Sinnerbrink: associate supervisor for the first six months of my candidature. I would like to thank Dr. Robert Sinnerbrink for his support, and for generously reading relevant section of the original German edition of Jaspers' *General Psychopathology*, and for helping me clarify Jaspers' concepts.

Prof. Jeanette Kennett: associate supervisor from the second year of my candidature to submission. I would like to thank Prof. Jeanette Kennett for her advice, her friendly support, and for her reassurance in my moments of doubt.

I would also like to express my appreciation to the following academics for their generosity in spending time advising me in an un-official capacity.

Dr. Lisa Bortolotti: I would like to thank Dr. Lisa Bortolotti for regularly meeting with me and my principal supervisor during her three months sabbatical at Macquarie University, during the first year of my candidature. Her advice in the early stages of my candidature was extremely valuable.

Assoc. Prof. Robyn Langdon: I would like to thank Assoc. Prof. Robyn Langdon for meeting with me and with my supervisors, and for offering invaluable advice on the cognitive science aspects of my thesis. I would also like to thank her for reading and commenting on several drafts of Chapter 6. She has continued to be encouraging and supportive, and always generous with her time.

Prof. Max Coltheart: I would like to thank Prof. Max Coltheart for inviting me to join the Australian Research Council of Excellence in Cognition and its Disorders (CCD) Belief Group at Macquarie University, and for his patience in answering my many questions.

I would also like to thank all members of the CCD Belief Group for their regular stimulating and informative discussions, without which this thesis could not have been written in its present form.

INTRODUCTION

1. Introduction

The study of delusions is founded in two different philosophical traditions: the continental or phenomenological tradition and the Anglo-American or analytic tradition. In this thesis I argue that the most plausible account of monothematic delusions must ultimately incorporate valuable insight from both traditions. I argue my case by presenting a new two-factor account of the Capgras delusion in which the characterization of each factor reflects insights from both traditions. In doing so, my intention is to demonstrate that an account reflecting insights from both traditions offers a better explanation of the characteristics of the Capgras delusion than those accounts which draw from only one tradition.

When bringing together insights from the two traditions, the historical differences in their method of inquiry need to be accommodated. Any brief account of these differences is likely to create something of a false dichotomy. Both approaches to investigating delusions cover a wide range of views in which the distinction between the two is not always as clear cut as the following brief account implies. Nevertheless, it is necessary to make some broad and general statements about the differences because successfully bridging the gap between the two traditions is an essential part of this thesis. The first main difference is that phenomenologists view delusion as a disturbed experience which may lead to a false belief, rather than as the false belief itself. The second is that they focus on understanding and gaining a proper description of the structure of the delusional experience rather than seeking to discover the biological, neurological or psychological cause of the false belief. The third is that they study whole persons embedded in their environment, rather than isolated cognitive processes.

Karl Jaspers, the founding father of psychopathology and advocate of the phenomenological method, published an account of delusion in 1913, which is still reflected in current research. "There is no doubt that Jaspers' influence lives on in modern Psychiatry" (Walker: 1988, 628). According to Jaspers, there is a "phenomenologically peculiar delusional experience", and his aim is to "find out what this basic primary delusional experience might be" (Jaspers: 1963, 97). Rather than attempting to define delusion, phenomenologists who follow Jaspers focus on describing their characteristics.

> As an approach, phenomenology is concerned with attaining an understanding and proper description of the structure of our experience; it does not provide a naturalistic explanation of it in terms of biological genesis, neurological basis, unconscious psychological motivation, or the like. (Parnas, Sass & Zahavi: 2008, 281)

For example, Gallagher says of his multiple realities account that it "does not provide a causal explanation of delusions, but aims to work out a more adequate characterization of delusions" (Gallagher: 2009a, 243). In contrast, analytic researches define delusions, relying heavily on the following oft quoted definition from the Diagnostic and Statistical Manual of Mental Disorders (DSM).

> A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. (DSM-IV: 2000, 281)

Notably absent from the definition is any reference to the patient's experience. According to Jaspers, the natural consequence of psychological inquiry that concerns itself only with objective data is "psychology without a psyche" (Jaspers: 1968, 1313). That this warning is still relevant today is evidenced by Lamme's suggestion that we move "our notion of mind towards that of the brain" (Lamme: 2006, 494) and by McDermott who envisages a future computational explanation of access consciousness in which phenomenology and experiences will 'drop out'. He suggests this will be the natural consequence of failure to find a neural locus for the 'self' and therefore, "appeals to 'what it is like' to experience such-and-such will carry little weight" (McDermott: 2007, 518). Not all analytic researchers reject experience, and some are moving towards a greater inclusion of experience in their accounts.

> We are seeing a reasonably broad-based attempt to bring phenomenological perspectives to bear on scientific psychiatry, particularly in schizophrenia, in the now dominant world of Anglophone psychiatry and clinical psychology. (Sass, Parnas & Zahavi: 2011, 2)

The above comments reflect the wide range of views currently being presented, and the blurring of the distinction between phenomenology and analytic research that is increasingly occurring.

When delusions are defined as beliefs, it follows that a delusion results from a disturbance somewhere in the formation and maintenance of the belief. The aim of most analytic researchers is to describe the causal processes that lead to each delusional belief. Cognitive processes are broadly conceived of as a series of discrete processes that transmit information, one information-processing system to another, readily represented by box-and-arrow charts. It is hypothesised that it is a disruption to one or more of the processes in the chain of processes that gives rise to a delusion. Therefore, the task of these analytic researchers is to isolate and describe the processes that are impaired, and to demonstrate how this could impact the processing chain in such a way as to produce a specific delusional belief. An hypothesis is formed regarding the site and/or the nature of the impairment which is thought to cause a particular delusion. The hypothesis is then subjected to empirical testing. The strength of the analytic approach is that the conception of cognitive processes as information-processing causal chains, lends itself to rigorous empirical testing. Recent advances in brain imaging techniques have given further depth to this approach. There is now a wealth of empirical data that directly relates to various types of delusion. But this strength is also weakness. Laboratory methodology, which as far as possible strips away everything that is not thought causally relevant to the delusion, creates an artificial situation that does not reflect the complex nature of the real world situation in which the delusion occurs.

For phenomenologists, delusions are a 'phenomenologically peculiar' *experience* that reflects a disturbance in the subject's experience of her world. It is true that the disturbed experience may give rise to bizarre beliefs, but it is held to be the disturbed experience that constitutes the delusion, not the process of belief formation that flows from the experience. The aim of phenomenologists is to

describe the characteristics of delusion and to give an account of the disturbance in the underlying experience that gives rise to those characteristics. The traditional method is phenomenological inquiry in which detailed reports of patients' experiences are compared, and their common features or characteristics noted. Hypotheses are formed regarding the nature of the disturbance in the underlying structure of experience that might give rise to the noted characteristics and features. The strengths and weaknesses of the phenomenological approach are the mirror image of those of the analytic approach. The phenomenological approach is inherently holistic. The disturbed delusional experience is not indicative of a single anomalous experience but represents the patient's general mode of existence and the way in which she relates to her world. The strength of the approach is that it maintains a global view of a patient and situates her in the world. But this strength is also weakness because whole experiences are difficult, if not impossible, to test empirically.

Historically, phenomenologists have focused their research on schizophrenia and the multiple-theme (polythematic) elaborated delusional belief webs, such as delusions of reference, grandeur and persecution, characteristic of the condition. As well as investigating schizophrenia, analytic researchers target circumscribed, single theme (monothematic) delusions, as these are more amenable to empirical testing. The differences between polythematic and monothematic delusions raise the question of whether or not the insights gleaned from investigating one are applicable to the other. At first glance, general deficit accounts of schizophrenia, such as those involving neurotransmitter dysregulation, do not appear applicable to monothematic delusions. Likewise, specific deficit accounts of monothematic delusions do not appear applicable to polythematic, elaborated delusions. More recently there have been attempts to bridge the divide. Gallagher's (2009a) multiple realities hypothesis accommodates both types of delusions, and based on Kapur (2003) dopamine dysregulation hypothesis of schizophrenia, Coltheart (2007) suggests delusions of reference could be explained in two-deficit (analytic) terms. Rather

than attempting to apply insights gained from analytic research into monothematic delusions to the delusional belief webs of schizophrenia, as does Coltheart, my aim is to introduce insights gained from the phenomenological study of schizophrenia into an account of a monothematic delusion. I use the Capgras delusion, the delusion that a loved-one has been replaced by an impostor, as my example. Mine is a hybrid account that combines insights from both traditions.

2. Thesis structure

The thesis is presented in two parts. Part One contains five chapters and Part Two, six chapters. In Part One, relevant literature on delusions is reviewed and critiqued. In Part Two, I offer a new account of both the first and the second factor in a two-factor account of the Capgras delusion. In the first chapter of Part One, three influential phenomenological researchers – Jaspers, Sass and Gallagher – are reviewed. Chapter 2 reviews and critiques Maher's (1988) 'explanationist' account, which challenges Jaspers' foundational view. It also describes various attempts to address the weakness in Maher's account. In Chapter 3 two-factor accounts of monothematic delusions are introduced. Two two-factor accounts, Coltheart's (2007) two-deficit account and Davies and Davies (2009) parametric variations account, are discussed and evaluated. In Chapter 4 two one-factor accounts, Gerrans' (2009) one-factor dopamine account and Corlett, Fletcher and colleagues' (2011) one-factor prediction error account are described and discussed. In the final chapter of the first section, two contentious issues relevant to the proposal I offer in the second section of the thesis are discussed. They are the relationship between experience and belief, and the contents of perception.

In Part Two I offer a new account of the first and second factor in a two-factor account of the Capgras delusion. In the first chapter I present a delayed-processing account of face encoding, which I propose constitutes the initial neuropsychological anomaly underlying the Capgras delusion. I describe how the anomaly gives rise to a data conflict, between 'familiar' and 'unfamiliar'

responses, which I argue is resolved by the 'impostor' attribution. In the second chapter I explain how the rich 'impostor' concept might be acquired if salient data from context is taken into account when making the attribution. The next chapter expands on this, proposing that the impostor concept is sourced from an alternative reality. The alternative reality is that of fiction. In the fourth chapter I suggest that if the Capgras patient's impostor is fictional, it explains unusual characteristics of the delusion such as double-bookkeeping, the bizarre content of the delusion, its incorrigibility and the fact that some patients are passive whilst others are violently aggressive in response to the delusion. In the fifth chapter I offer a coherence-monitoring account of the second factor in the Capgras delusion. This involves the decoupling of a conflict monitoring system from executive control, which is automatically stimulated to allow the (fictional) impostor attribution to pass unchallenged. The final chapter briefly describes how the two factors described in the previous chapters might be applied to several other monothematic delusions and to the delusions characteristic of schizophrenia.

In order to make clear the overall shape of my argument, I provide below brief descriptions of the contents of each chapter.

3. Part One

Chapter 1: Phenomenology and delusions

i) Karl Jaspers

According to Walker, "Jaspers is absolutely central to any discussion of the nature of delusions" (Walker: 1991, 94). Jaspers was the first to question the basic concepts of psychiatry, such as disturbed perception and thought. He argued for the view that there are individual mental symptoms, and shifted the emphasis in research from seeking an understanding of the individual patient to the understanding of general principles. It was Jaspers who described delusions as judgments that provide the meaning component of the immediate perceptual experience, clearly distinguishing them from other phenomena of mental illness,

such as hallucinations (Spitzer: 1990). Jaspers' account of delusions incorporates several specific concepts. I begin my review of his account by clarifying these concepts, explaining what he means by the terms psyche, understanding, form and content, and reality. I then give an account of his primary and secondary delusions, and of his conception of primary delusion as being a disturbance in the meaning component of the immediate perceptual experience. I note that a weakness in Jaspers' account generally is held to be his contention that primary delusions are beyond our understanding.

Jaspers' account of delusion as a disturbance in meaning is basic to my account of the Capgras delusion. I hypothesise that the content of the delusional belief (impostor) reflects the attributions of meaning to anomalous data produced by impaired face encoding. Thus the delusion reflects a disturbance in the meaning component of the immediate perceptual experience.

ii) Louis Sass

Jaspers' followers generally reject his pessimistic view that primary delusions are beyond our understanding. Sass offers the following explanation.

> No less a mind than Karl Jaspers believed, in fact, that any attempt at unriddling the enigma of schizophrenic consciousness was doomed to failure, and that we ought simply to acknowledge a fundamental unknowability, perhaps restrict ourselves to the investigation of underlying neurobiological causes. But there would be certain dangers in adopting this attitude of interpretive nihilism, for it risks doing a double disservice: first, to the patient, who would thereby be banished from the community of human understanding; and second, to the rest of us, who would be deprived of all access to what may be an important limit-case of the human condition. (Sass: 1992, 97)

Phenomenological researchers such as Sass attempt to give an account of the underlying disturbance that alters the patient's way of relating to her world. According to Sass, delusions characteristic of schizophrenia reflect a fundamental disturbance in ipseity. He describes the disturbance as having two aspects; diminished self-affection and hyperreflexivity. In this chapter I explain these two aspects of ipseity disturbance, and their relationship to each other. I then describe how Sass's conception of body alienation, unworlding and subjectivization are involved.

My own account of the Capgras delusion does not draw directly from Sass's account. However, it is influenced by several of his ideas. These mostly relate to his description of subjectivization, more specifically, his idea that delusion reflects a world of inner fantasy which has developed such a degree of 'phantom concreteness' that it leads to a quasi-reality, supported by belief-like states.

ii) Shaun Gallagher

Gallagher also rejects Jaspers' pessimism. He develops Sass's conception of delusion as an alternative reality, and offers a multiple realities hypothesis as a framework for understanding delusions. Before describing his hypothesis, I explain his conception of cognition as embodied and embedded, and how he differentiates between the body schema and the body image. Gallagher's multiple realities hypothesis is based on our normal experience of moving between everyday reality and the alternative realities of dreams, imagination, virtual world, and those of the theatre and cinema. He argues that entering these alternative realities requires an existential shift because it involves a different way of *being-in-the-world.* He suggests that we may be able to enter a delusional reality in the same way as we enter other alternative realities.

My account draws heavily on Gallagher's hypothesis in that I suggest, as a result of the second factor, the patient automatically, and involuntarily enters an alternative reality. But rather than being a separate delusional reality, I argue that the patient enters a fictional world.

Chapter 2: Brendan Maher

Brendan Maher is also essential to any discussion of delusions because he introduced a completely new way of conceiving of delusions that differed radically from Jaspers' account. Rather than describing the problem as a disturbance in the thought processes that form judgments about the encoded sense data, Maher described delusions as explanations of an anomalous experience caused by a neuropsychological anomaly with a biological basis. He argues further, that when seeking an explanation, the patient's thinking is

perfectly normal. Thus, he makes two claims. The first is that there is an underlying impairment that gives rise to an anomalous experience. The second is that the patient's thinking is normal. The chapter is divided into two parts. The first describes and discusses Maher's account. The second described various attempts to verify or falsify his two claims, and to address the weaknesses in his account.

The first section begins with an account of Maher's (1999) challenge to Jaspers. It is followed by an explication of Maher's (1988a) claim that an anomalous experience with a biopathological basis underlies all delusions, and that the patient's explanation of the anomalous experiences is rational. This is followed by four criticisms of Maher's account and his responses to those criticisms. The second section begins with the various attempts to verify Maher's two claims. It describes the empirical data that supports the first claim. Attempts have been made to identify abnormalities in the thinking processes of delusional patients, and thus prove, or disprove, Maher's second claim. Possible factors, such as conservation and observational adequacy, attributional bias and probabilistic reasoning bias, are discussed in this chapter. The results of research into the second claim remain controversial. However, as some subjects have the anomalous experience but not the delusion, it is generally considered that Maher's account provides insufficient explanation of the formation and maintenance of a delusion. A second factor is needed.

In keeping with Maher's account, the first factor in my account is impairment to a cognitive process. I hypothesise that the first factor in the Capgras delusion is a delay in face encoding. However, in my account, this does not lead to an anomalous experience, which requires explanation, but to anomalous data that is accommodated by the 'impostor' concept.

Chapter 3: Two-Factor Accounts

Two-factor accounts hold that two factors are necessary for the formation and maintenance of a monothematic delusion. In this chapter I describe two two-

factor accounts, one offered by Coltheart (2007) and the other by Davies and Davies (2009). A 'factor' refers to any cause of the formation and maintenance of a delusion, whereas a 'deficit' refers to a factor that is thought to be an impairment, or loss of, some aspect of neuropsychological function. Coltheart's account is a two-deficit account. It specifically applies to a group of monothematic delusions for which a possible first factor has been identified, although Coltheart anticipates that future research will see it applied to all monothematic delusions. In his account, the first deficit is a neuropsychological impairment that gives rise to anomalous data, with an impairment being specific to a particular delusion. The second deficit is impairment to a putative belief evaluation system, and this is hypothesised to be common to all delusions. A description and discussion of his account and its application to the Capgras delusions is given in this chapter.

Davies and Davies offer a parametric variation account of monothematic delusions. They accept a neuropsychological deficit and a possible accompanying anomalous experience as the first factor. The first variation they note is that the anomalous experience may be *explained* or *endorsed*. They then suggest that both explanation and endorsement are subject to further parametric variations, such as the degree to which the content is encoded, and whether the explanatory hypothesis is normal, biased or impaired. Like Coltheart, they suggest the second factor is impairment to a putative belief evaluation system, but they cash this out in terms of dual processing accounts of reasoning, judging and decision making. It enables them to make the more specific claim, that the second factor is impairment to working memory, or to executive function, with a neural basis in damage to the right frontal region of the brain.

My account is a two-factor account, not a two-deficit account. I suggest the second factor is a normal process that is applied to an abnormal situation. Further, when the second factor is present, it is stimulated by the first factor, rather than being a separate factor that co-exists with the first.

Chapter 4: One-Factor Accounts

Despite the popularity of two-factor accounts in the analytic literature, some researchers suggest it is possible to give a plausible one-factor account of delusions. In this chapter I describe two such accounts. The first is Gerrans' (2009) one-factor account based on dopamine dysregulation. The second is a prediction error account offered by Corlett, Fletcher and colleagues (2011). Gerrans' account explains how a dopamine dysregulation affects the processes involved in constructing an autobiographical response to experience. There are evolutionary, neurobiological and phenomenological considerations that underlie his account, and I begin by describing these. His explanation of delusions involves the employment of offline resources needed when dealing with novel situations, and the increasing decontextualization that enables further offline resources to be employed when necessary. Dopamine is implicated in this process, and according to Gerrans, dopamine dysregulation renders the patient unable to access the offline resources necessary to resolve an anomalous situation. I give a description of the normal and abnormal processes that constitute Gerrans account. In addition, I acknowledge that more recently, Gerrans (2012) has offered a two-factor account.

Corlett, Fletcher and colleagues propose a one-factor account involving a prediction error Bayesian model of delusions, which they claim subsumes the standard two-factor account into a single factor, an aberration in Bayesian inference. Theirs is a complex account which attempts to integrate prediction error theory, Bayesian probabilistic reasoning and neuropharmacology. They give a brief explanation of the Capgras delusion in terms of their account. I describe the various elements in their account, and their explanation of the Capgras delusion. I then suggest that their account is not, in fact, a one-factor account, arguing that it is illegitimate to subsume two causal factors into a single non-causal factor.

Chapter 5: Two contentious issues

There are two contentious issues important to any discussion of delusions. The first is the doxastic status of delusions and the degree to which experience determines the content of the belief. The second regards the richness or poverty of the perceptual content. In the first section of the chapter I discuss the doxastic status of delusions. I begin with an acknowledgment of three non-doxastic accounts – those of Berrios (1991), Currie (2000), and Sass (1992). In keeping with the definition given in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR), most analytic researchers regard delusions as beliefs. However, the way in which the belief is thought to be formed is a matter of debate. Supporters of 'endorsement' accounts are influenced by the phenomenologists' contention that delusions are disturbed experiences. They hold that the belief is formed by taking the experience as veridical. In contrast, explanationists follow Maher, and hold that the content of the belief is an explanation of an anomalous experience, or, in some accounts, of anomalous data. I discuss the two views, their strengths and weaknesses, and the challenges each approach faces. I then describe two recent accounts that attempt to combine the strengths of both endorsement and explanationist accounts, whilst ameliorating their weaknesses. The account of the Capgras delusion I offer in this thesis is an endorsement account in that I argue that the patient experiences her loved-one as an impostor, and that this experience is taken as veridical.

The second section is devoted to discussing the issues that arise when attempting to define perception and its contents. I discuss three different views—those of Fodor (1983), Churchland (1988) and Raftopoulos (2001). These three researchers focus on sub-personal processes and conceive of perceptual content as being relatively thin. Gallagher takes a different approach, focusing on the subject's subjective experience, rather than on the sub-personal processes. I describe his more liberal interpretation of the term 'perception', which includes rich contents as part of the immediate perceptual experience. The direct nature of perception and the embedded, embodied and enactive

nature of cognition are important to his account. In my account of the Capgras delusion I argue for rich perceptual content, drawing from Gallagher's account.

4. Part Two

Chapter 6: The Capgras Delusion: a delayed-processing account of factor 1.

In the first chapter of this section, I offer a new account of the first factor in a two-factor account of the Capgras decision. In keeping with the analytic tradition, I hypothesise that the first factor is a neurological impairment. I suggest that it is a delay in the structural encoding of the face, which causes a familiar face to be erroneously classified as unfamiliar. A fuller explanation of the relevance of the speed of encoding will be given later (Chap 6: *3)*. I argue that this sets up a *prima facie* expectation that the person is unknown, an expectation that is violated when the person's identity is correctly identified. My account draws on Whittlesea and William's (2000) discrepancy attribution theory of the source of the subjective feeling of familiarity. According to their account, the fluency of the sub-personal processing of any given sense data is monitored and compared with expectations that are based on prior experience. As known faces are processed more rapidly than unknown faces, a slowing of face encoding of a loved-one's face will result in the face being classified as unfamiliar, based on prior experience.

My account follows the Ellis (2007) face processing model, which is reproduced in this chapter. However, rather than locating the impairment after the face has been recognized, but before the person's identity has been accessed, as Ellis does, I locate it at the site of the initial encoding of the raw face data. In doing so, I argue, it becomes easier to account for the anomalous skin conductance responses (SCR) exhibited by Capgras patients, the tendency of the delusion to spread over time, and for Capgras delusions about inanimate objects.

Chapter 7: Meaning and the Capgras delusion

In this chapter I explain how the Capgras patient's immediate perception of her loved-one might acquire the rich 'impostor' concept. I begin by noting that our

conception of the thickness or thinness of perceptual content can be biased by both over-describing and under-describing the immediate experience, and I give an example for each drawn from the literature on delusions. I then argue that we see objects embedded in context. I discuss empirical data that demonstrates that we grasp the gist of a scene in milli-seconds, and argue that the demands of survival suggest that we also grasp the gist of the meaning of the scene with equal rapidity. Using Coltheart's account as my example, I argue that the explanationist account of the Capgras delusion is inadequate, because the attributing of meaning is based solely on the output of face processing, and other salient data is not taken into account.

If, as I argue, the face of the loved-one is perceived as being embodied and embedded, and the meaning of the immediate perceptual experience encompasses the gist of the scenario, the process must include an interpretation of the intentions of the 'stranger'. I give a brief description of the two main theories, Theory Theory (TT) and Simulation Theory (ST), which describe the mechanism by which we understand another person's mental states. I give a more detailed account of a theory offered by Gallagher, and note my leanings towards his account. I then explain how the interpretation of the anomalous data familiar/unfamiliar provided by the delayed processing of the loved-one's face, when interpreted in context as part of the gist of the meaning of the scene, results in the 'impostor' attribution.

Chapter 8: Alternative realities and delusions

In this chapter I argue that the 'impostor' concept that provides the content of the Capgras delusion is drawn from the patient's conception of an alternative reality. As noted in previous chapters, Jaspers suggests that delusion involves an alternative reality when he describes primary delusions as reflecting 'a world of new meaning'. Sass describes them as reflecting an 'inner world of fantasy', and Gallagher suggests it is possible that patients enter a delusional reality in much the same way as we enter any other alternative reality. My account is in keeping with these views, more specifically with that of Gallagher. I hypothesise that, as

a result of the second factor, the Capgras patient automatically and involuntarily enters the alternative reality of fiction.

If delusions involve alternative realities, it raises issues concerning the way in which we create alternative realities and engage in imaginary worlds. I describe one of the many explanations of the process – that offered by Walton (1990), primarily in his book *Mimesis and Make-believe.* I then argue that, for most people, the 'impostor' concept is formed from engaging in fictional narratives. Consequently, the neural assemblies connecting the concept to an alternative reality will be dominant. When the rapid, non-reflective process select concepts to provide the meaning component of the immediate perceptual experience, the dominant concepts are automatically selected. Thus, I argue, the impostor concept linked to the web of beliefs associated with an alternative reality will be automatically selected.

Chapter 9: Explaining unusual characteristics

In this chapter I argue that if alternative realities are involved in delusions, it helps to explain some of the puzzling characteristics of the Capgras delusion, thus giving further support to the hypothesis. The first is the fact that whilst some patients exhibit a puzzling passivity in response to their delusional belief that their loved-one has been replaced by an impostor, others respond aggressively. I argue that passivity reflects the way in which we interact with fictional characters, such as those portrayed by actors at the theatre, or in a novel. However, the degree to which we become absorbed in a fiction varies, and as our absorption increases, so does our emotional engagement. I argue that aggressive patients are those who have become totally caught up in the story, but like all patients, they are unaware they are engaged in a fictional reality. A possible criticism is that it is unlikely that patients would fail to notice the shift from everyday reality to a fictional reality when viewing their loved-one. The experiences are phenomenologically different, and normally we are able to distinguish between worlds. In response I argue that we are not always aware of the reality in which we are engaged, and describe the Stanford Prison

Experiment as a graphic example of the ease with which subjects can fail to maintain an awareness of the reality from which they are operating.

If delusions involve alternative realities, it helps to explain two other puzzling characteristics. The first is double-bookkeeping, which is the patient's ability to firmly hold a belief whilst at the same time behaving as if the belief were untrue. The second is circumscription, which is the fact that the delusional belief does not impact upon, nor appear to be influenced by, the patient's general web of beliefs. Fictional worlds are constituted by a sub-set of beliefs separated from the belief web of the everyday world. Therefore, if the content of the delusion belongs to a fictional world it might explain both of these characteristics. It might also help explain the incorrigibility of the delusion. As fictional worlds operate according to a different set of rules, values and logical possibilities to that of everyday reality, normal rational arguments may simply not apply. Further, the fictional world hypothesis could explain the bizarre content of the delusion, and also, the patient's high tolerance for the unexplained. Patients rarely know why a stranger is impersonating their loved-one, or what has happened to the loved-one, yet it often fails to bother them. This might reflect an engagement in a fictional world, because we accept that fictional worlds are incomplete and we have to wait for the story to unfold before we are able to understand why certain events and situations have arisen.

Chapter 10: Factor 2: A coherence-monitoring account

According to my account, as I have indicated in previous chapters, the patient enters an alternative reality on viewing her loved-one. In this chapter I describe how the second factor causes this to occur. I begin by revisiting other accounts of the second factor, and Whittlesea and William's description of process monitoring. I then discuss our normal ability to keep alternative realities apart. Drawing on Gallagher's contention that engaging in alternative realities involves an existential shift, I suggest that the information linked to a concept might include an existential 'tag' denoting the world to which the concept predominately belongs. I then discuss the alternative reality of hypnosis. I

describe two theories of hypnosis, social interaction theory and dissociation theory, from which I draw in my account. I give an account of empirical data that indicates that entry into the hypnotic condition involves the decoupling of a conflict monitoring system from executive control. I then argue that the decoupling is a normal mechanism that has evolved to enable entry into imaginative worlds. I apply this to the Capgras delusion in the following way. When the (fictional) impostor concept is accessed to resolve the conflict (familiar/unfamiliar) created by the first factor, it creates a second conflict. Applying a concept tagged as fictional to a person in the real world creates a cross-world conflict. To resolve the second conflict, the monitoring system automatically and involuntarily decouples from executive control to allow the attribution to pass unchallenged. Thus, a normal mechanism is brought to bear on an abnormal situation. The automatic, involuntary decoupling of the conflict monitoring system from executive control constitutes the second factor. According to my hypothesis, the first and second factor produce the first-pass, non-reflective interpretation of the gist of the scenario that results in the immediate perceptual experience.

Chapter 11: Further applications

In this chapter I briefly explain how my hypothesis might be applied to five other monothematic delusions. I also make the very broad suggestion that it may be applicable to schizophrenia. I suggest the first three monothematic delusions I describe have the same putative first and second factor as the Capgras delusion. That is, the first factor is a processing delay, which produces false data, resulting in a meaning attribution linked to fiction, and the second factor is the automatic, involuntary decoupling of a monitoring system from executive control to resolve the cross-world conflict. The three delusions are, mirrored-self misidentification without mirror agnosia (my reflected image is another person), thought insertion (someone, or something, is inserting thoughts into my mind), and thought control, (someone, or something, is controlling the thoughts in my mind).

The next two delusions I discuss also involve an alternative reality, but not that of fiction. They involve the different realities we experience at different stages of cognitive development, and they reflect the different developmental concepts the subject has acquired at each level. At any stage in our development, impairment can result in cognitive regression and the partial return to earlier patterns of reacting and thinking. Two delusions that clearly involve such regression are mirrored -self misidentification with mirror agnosia (loss of knowledge of how mirrors work) and somatoparaphrenia, which is the belief that a body part, usually one's limb, belongs to another person. The latter involves the loss of the concepts of body form and integrity. Although the first factor differs, in that the underlying impairment causes cognitive regression rather than a processing delay, I suggest that when the patient views one aspect of her world from a different conceptual perspective from that of the rest of her world, it creates incoherence. If we take the different cognitive developmental levels to be different realities, the inserting of a child perspective into the normal adult perspective creates cross-world confusion and should attract the attention of the conflict monitoring system. I suggest that the second factor remains the same—a decoupling of the monitoring system from executive control to allow the anomaly to pass unchallenged.

Although schizophrenia is not the subject of this thesis, I conclude with very broad suggestions regarding the application of my hypothesis to this condition. I begin by noting that schizophrenia describes a cluster of symptoms, and therefore, it is unlikely that there is one underlying cause. I note that auditory hallucination, which is one of the characteristic symptoms of schizophrenia, involves an endogenous activation of the auditory cortex. Thus, the experience could be explained using my two-factor hypothesis. I then note that the bizarre content of the delusions associated with schizophrenia has led researchers, such as Jaspers, Sass and Currie, to suggest they reflect an alternative reality. If this is the case, the decoupling of the monitoring system from executive control would be involved. Two brain areas implicated in this process are the anterior cingulate cortex (conflict monitoring) and the dorsolateral prefrontal cortex

(executive control). Empirical data show that patients suffering schizophrenia have impairment at both these sites. For example, increased metabolic activity in the right anterior cingulate of patients during conflict tasks has been observed (Nordahl et al.: 2001), and "Schizophrenia has been consistently shown to be associated with impairment to executive function" (Westernhausen, Kompus & Hugdahl: 2011, 172). I make the broad suggestion that general impairment of conflict monitoring and/or executive control could result in all manner of bizarre cross-world meanings becoming attached to the immediate perceptual experience, and allow them to pass unchallenged. As a result, patients might experience a confusion of meanings drawn from religious and childhood perspectives and the multiple worlds of the imagination.

PART ONE

"There are in psychopathology a number of viewpoints, a number of parallel approaches which in themselves are quite justifiable and complement rather than oppose each other."

Karl Jaspers *General Psychopathology* 1959 Eng. Trans 1963
Baltimore: John Hopkins University Press, 1997, p xviii

"Some delusions do appear to be meaningless and incomprehensible, but the default presumptions must be that they make sense."

Jennifer Radden *On Delusions*
London and New York: Routledge, 2011, p xii

CHAPTER 1: PHENOMENOLOGY AND DELUSIONS

1. Introduction

Phenomenology is both a philosophy and a method and has its origins in the works of Edmund Husserl (1859-1938). The phenomenological method was introduced into psychiatry by Karl Jaspers, principally in his work *General Psychopathology* (1913). Phenomenology attempts to describe our subjective, first-person experience directly, and independently of any causal explanation that might be given. For Husserl it is an *a priori* investigation of the essence or meaning common to the thoughts of different minds. Heidegger, Sartre and Merleau-Ponty, whilst not necessarily accepting Husserl's conclusions, pursued and refined the phenomenological method. The three investigators of mental illness included in this chapter—Jaspers, Sass and Gallagher—base their theories on a phenomenological world view drawn from one or more of the above philosophers. Phenomenologists predominantly discuss delusions in terms of schizophrenia and the polythematic, elaborated belief webs characteristic of that condition. They rarely discuss the monothematic (single theme), circumscribed delusions that have dominated more recent analytic research. It is my intention in this thesis to take selected insights provided by the three phenomenologists discussed in this chapter and to incorporate them into a new account of the Capgras delusion, and further, to suggest briefly how the account might apply to five other monothematic delusions.

Section One: Karl Jaspers

1. Introduction

It is generally agreed that the work of Karl Jaspers represents a turning point in the investigation of mental illness. Jaspers was the first to question the basic concepts of psychiatry and to address the question of how symptoms in the mental realm can be established. He challenged the validity of the diagnostic

criteria for delusions and introduced the idea that form (the way a belief is held) is more important than content when it comes to classifying and identifying delusions.

> Many authors have pointed out that the beginning of scientific psychopathology is marked by the appearance of Karl Jaspers' *General Psychopathology* in 1913. (Spitzer: 1990, 5)

He was the "first psychiatrist to clearly identify the need to be clear about one's methods" (Ghaemi: 2007, 79), and he "laid the foundation for the scientifically grounded diagnostics and nosology which have permeated modern classification systems" (Bürgy: 2008, 1200).

> Arguably it was Karl Jaspers of all those influenced by the phenomenological movement who had the greatest influence on psychiatry or at least mainstream psychiatry in the English-speaking world. (Mullen: 2007, 116)

Jaspers' influence can be seen in the Diagnostic and Statistical Manual of Mental Disorders (DSM). For example, the DSM definition of hallucination reflects his recognition that hallucination is like perception. The DSM definition of delusions also draws on his work.

> As with hallucination, the DSM-III-R definition of delusions can easily be traced back to Jaspers, whose famous "criteria" of delusions – certainty, incorrigibility, and falsity – are mentioned in its very first sentence. (Spitzer: 1990, 6)

Jaspers' account of delusions is quite complex. Therefore, I will begin by clarifying some of the important concepts on which it is based.

2. Concepts

i) The psyche

One way to interpret Jaspers' concept of the psyche is to say that it is the underlying organising principle that makes a homo sapien a human *being.* The term *being* implies the ongoing moment to moment flow of existence, the continuous unfolding of the individual which always takes place within the world. According to Jaspers there is no valid theory of the psyche, "only a philosophy of human existence**"** (Jaspers: 1963, 551). The psyche is not an object but simply "*the encompassing of existence"* (Jaspers: 1963, 9). In clarification he offers the following descriptions.

> Psyche means *consciousness,* but just as much and from certain points of view it can even, in particular, mean 'the unconscious'. (2) Psyche is not to be regarded as an object with given

> qualities but as '*being in one's world*', the integrating of an inner and outer world. (3) Psyche is a becoming, an unfolding and a differentiating, it is nothing final nor is it ever fully accomplished. (Jaspers: 1963, 9)

> *Psyche* means direct inner experience (the material for phenomenology); it means whatever it is that produces meaningful function or appears as human expressiveness; it means the unity of the 'I', the self, the underlying psychic substance, etc.. (Jaspers: 1963, 222)

Despite the distinction we can make between body and psyche, he notes in every human being we see the unity of body and psyche appearing as a living whole. Jaspers rejects Cartesian dualism, monistic materialism and spiritualism in favour of "*the one absolute* of the substantial Being of body-psyche unity" (Jaspers: 1963, 224). However, although the body-psyche is one for us in expression, we are unable to grasp the transcendental whole.

> The completeness of the whole can never be encompassed by the time-bound nature of human knowledge. Knowledge is only true within that part of space accessible to us. If we want to know the transcendental whole, which has both physical and psychic effects and is primarily both, we find it has vanished from us into the clarity of particular facts which are comprehensible but which are never the whole itself. (Jaspers: 1963, 225)

We can never grasp the totality of Being because it is a constant 'becoming' or 'unfolding'. It is not a static object. We can only gasp particular facts as they are presented to us in time and space. This creates a further problem when attempting to comprehend the transcendental whole.

> The act of knowing puts us in possession of all Being only *as this falls apart into subject and object.* That is, we possess it as an object for our awareness, not as it is in itself but as it seems to be for 'consciousness in general' in this divided condition. (Jaspers: 1963, 757)

The transcendental whole transcends the subject/object dichotomy, but it is inaccessible to us because we are only able to possess knowledge in the divided form of subject and object. "The whole is an idea only" (Jaspers: 1963, 756).

Jaspers says psychopathologists aim "to know, recognise, describe and analyse general principles rather than particular individuals" (Jaspers: 1963, 1). The subject-matter is "actual conscious psychic events" (Jaspers: 1963, 2). However, he assumes that, although not always evident, all abnormal psychic events are determined by abnormal neurological activity.

> In many psychic disturbances and personality disorders (psychopathies) we find nothing at all in the brain which could offer a direct or more remote basis for the event. In spite of this it can hardly be doubted that every peculiar psychic event will also have its own peculiar somatic determinants. (Jaspers: 1963, 458-459)

Although he notes the lack of correspondence between "clearly *analysed psychic function* and anatomically finely divided areas" (Jaspers: 1963, 492), he says "it is beyond doubt that there is a *relation between cerebral illness and psychosis"* (Jaspers: 1963, 492). He predicts correctly that the issue of localisation will be clarified as a result of further research into the structure of the brain. This does not mean Jaspers believes abnormalities such as delusion are always caused by organic disease. He says, for example, in schizophrenia "there is no underlying destructive organic process nor is the "machinery smashed up" (Jaspers: 1963, 608). He repeatedly stresses the reciprocity of the individual's inner and outer world and of the body and the psyche, wherein each shapes and, is in turn, shaped by the other. "One way causality … does not exhaust life's possibilities" (Jaspers: 1963, 452). As Jaspers views the individual as a 'whole' there will be many factors impinging on the formation of abnormal psychic states.

ii) Understanding

Jaspers differentiates between the *manifestations* of psychic events and the psychic event itself. If, for example, the content of the psychic event is 'fear', it may or may not exhibit 'somatic accompaniments' such as increased pulse rate and the like. There are two ways we can pursue understanding other people's experiences, 'objectively' and 'subjectively'. Jaspers defines objective as "everything perceived by the senses" and subjective as everything that can be comprehended by "*empathy* into psychic events, or some realisation of psychic contents" (Jaspers: 1963, 26). Having drawn these distinctions, he notes that empathetic understanding may be accessed through the objective. The outward signs of some psychic content, such as signs of fear, can be directly understood empathetically[1] (Jaspers: 1963, 26). Jaspers also describes two ways in which

[1] Jaspers rejects the notion that our understanding of expression rests on conclusions drawn from analogy from one's own psychic life and applied to that of others because, he claims, we understand an expression directly, without any need for reflection. He notes "We understand in a lightning flash at the very moment of perception" (Jaspers: 1963, 254). However, this immediacy does not guarantee our understanding is valid. Deceptions are numerous. Reasoning comes afterwards as a strategy for controlling error.

we can explore connections—by the perception of causal connections and by the perception of meaning. He calls the former *explanation* reserving *understanding* for the latter. Understanding itself has two different meanings, *static* and *genetic*. The understanding of a psychic state or quality is termed *static.* The understanding of the connection between the states and qualities, how one emerges from the other, is termed *genetic.*

> For instance, thoughts may be understandable because they emerge from each other according to the rules of logic and here the connections are understood rationally (that is, we understand what is said). But where we understand how certain thoughts arise from moods, wishes or fears, we are understanding the connections in the true psychological sense, that is by empathy (we understand the speaker). Rational understanding always leads to a statement that the psychic content was simply a rational connection, understandable without the help of any psychology. Empathetic understanding, on the other hand, always leads directly into the psychic connection itself. Rational understanding is merely an aid to psychology, empathetic understanding brings us to psychology itself." (Jaspers: 1963, 304)

iii) Form and content

According to Jaspers, psychic events consist of *form* and *content*. All experience entails both incoming sensations (content) and an organising concept (form).

> Form must be kept distinct from content which may change from time to time, e.g. the fact of a hallucination is to be distinguished from its content, whether this is a man or a tree, threatening figures or peaceful landscapes. Perceptions, ideas, judgments, feelings, drives, self-awareness, are all forms of psychic phenomena; they all denote the particular mode of existence in which content is presented to us. (Jaspers: 1963, 58-59)

Jaspers claims, that in cases of mental illness, classification by content is unsuccessful because of its transitory and often arbitrary nature, and because the same content can appear in any form. Also, content provides little insight into the nature of the underlying disturbance that gives rise to psychic events. What is important to determine is whether the psychic event is a disturbance in perception, for example, or ideas, judgment and the like. But he points out there is a connection between content and psychosis. Schizophrenia transforms the content of experience*.* It is not merely chance contents of general humankind haphazardly interpreted into meaningless structures (Jaspers: 1963, 282). For Jaspers, understanding delusion requires that we determine the form or mode of existence in which content is presented, and this is achieved through phenomenological enquiry.

> Phenomenology only makes known to us the different forms in which all our experiences, all psychic reality, take place: it does not teach us anything about the contents of the personal experience of the individual. (Jaspers: 1968, 1323)

And Jaspers notes, from a phenomenological point of view, it is only form that is of interest.

iv) The issue of reality

According to Jaspers', "the experience within which delusion takes place is that of experiencing and thinking that something is real" (Jaspers: 1963, 93). He defines three characteristics of the reality we experience. First, what is real is what we *concretely perceive* as opposed to what we imagine. Hallucination reflects disturbed perception because the content of the hallucination, which only exists in the patient's mind, is perceived as having concrete existence in the external world. Second, reality lies in the simple *awareness of Being.* The primary experience of our self as 'real' is our basis for accepting the reality of the external world. But it can be lost in derealisation and depersonalisation, such as in the Cotard delusion, which involves the belief one is dead. Third, what is real is *what resists us*. What resists us is anything that inhibits our bodily movement or prevents the immediate realisation of our aims or wishes.

As well as the above, Jaspers notes there is also an objective world, a general world common to all. Participation in this insures a criterion for accuracy of thinking and its objective validity (Jaspers: 1963, 12-13).

> Normal convictions are formed in a context of social living and common knowledge. Immediate experience of reality survives only if it can fit into the frame of what is socially valid or can be critically tested. (Jaspers: 1963, 104)

And for Jaspers, the reality we experience is a primary (irreducible) phenomenon:

> But the reality we meet in practice is always an *interpretations*, a meaning, the meaning of things, events or situations. When I grasp the meaning, I grasp reality. (Jaspers: 1963, 94)

According to Jaspers, it is a disturbance in the meaning of the patient's experience of reality that is the primary delusion experience. "The awareness of meaning undergoes a radical transformation" (Jaspers: 1963, 99). Further, patients seem unable to correct their false judgments of reality either in reference to their other beliefs, in light of the common experience of members of their social and cultural group, or in the face of objective reality. The

relationship between meaning and reality is central to Jaspers' account of delusion.

3. Delusions

Jaspers begins by making an important distinction between two forms of delusions according to their origins.

> We can then distinguish two large groups of delusions according to their *origin*: one group *emerges understandably* from preceding affects, from shattering, mortifying, guilt-provoking or other such experiences, from false-perception or from the experience of derealisation in states of altered consciousness etc. The other group for us is *psychologically irreducible*; phenomenologically it is something final. (Jaspers: 1963, 96)

He gives the term *delusion-like ideas* to the former and *delusion proper* to the latter. With *delusion-like ideas* there is only a tendency towards a false judgment with a transitory certainty that may be corrected. With *delusion proper* all doubt has ceased. The patient holds to the certainty of the object's reality in spite of the known objections and in spite of reflection, overcoming any initial doubts she may have had. Having drawn this clear distinction, he then suggests the term delusion should "properly only be given to those delusions which go back to primary pathological experiences as their source" (Jaspers: 1963, 106). The term delusion-like ideas should be reserved for those 'so-called delusions' that emerge understandably from other psychic events or can be traced back to drives, desires and fears. The distinction is important because it is generally accepted that Jaspers describes delusional beliefs as being qualitatively different from normal beliefs. This is only true when 'delusion' is referring to 'delusion proper', because these delusions are beyond our understanding. The other broad group of delusions (delusion-like ideas) can be understood in terms of drives, desires and fears.

Jaspers' first question regarding delusion proper is "what is the primary experience traceable to the illness?" (Jaspers: 1963, 96). This question is not easily answered.

> If we try to get some closer understanding of these primary experiences of delusion, we soon find we cannot really appreciate these quite alien modes of experience. They remain largely incomprehensible, unreal and beyond our understanding. Yet some attempts have to be made. We find that there arise in the patient certain primary sensations, vital feelings, moods,

> awarenesses … Patients feel uncanny and that there is something suspicious afoot. Everything gets a *new meaning.* The environment is somehow different. (Jaspers: 1963, 98)

He describes the uncanny feeling as the *delusional atmosphere.* However, he notes this analysis does not hold in all cases because sometimes content is immediately present and vividly clear. Delusional perception, for example, may range from an experience of "some vague meaning to clear, delusional observations" (Jaspers: 1963, 99). There is some tension between his description of delusion proper as "the vague crystallisations of blurred delusional experiences and diffuse, perplexing self-reference" (Jaspers: 1963: 107) and his acknowledgement that some delusions have clear and vivid content. This tension seems best resolved by interpreting Jaspers as saying that the psychic disturbance gives rise to a diffuse pervasive feeling of oddness (delusional atmosphere), which in specific situations crystallises into delusional content that ranges from vague to vividly clear. Having described the delusional atmosphere, Jaspers explores the original experience further, not by studying vague moods and feelings, or the content which is often accidental, but by exploring the "psychological significance … of this delusional experience of reality in which the environment offers a world of *new meanings*" (Jaspers: 1963, 99). It is here we come to Jaspers' central idea that the primary delusional experience involves a disturbance in the meaning component of the patient's immediate experience.

Jaspers says "all thinking is thinking about meaning" (Jaspers: 1963, 99). Further, if the meaning is perceived directly by the senses or is directly present in imagination or memory, "the meaning has the character of reality" (Jaspers: 1963, 99). He begins his discussion of disturbed meaning with an account of normal thinking and the meaning that accompanies immediate perception.

> Perceptions are never mechanical responses to sense stimuli; there is always at the same time a perception of meaning. A house is there for people to inhabit; people in the streets are following their own pursuits. If I see a knife, I see a tool for cutting. If I look at an unfamiliar tool from another culture, I may not see its precise meaning but I can appreciate it as a meaningfully shaped object. We may not be explicitly conscious of the interpretations we make when we perceive but nevertheless they are always present. Now, the *experiences of primary delusion are analogous to this seeing of meaning,* but the awareness of meaning undergoes a transformation. (Jaspers: 1963, 99)

In further clarification he says "there is an immediate, intrusive knowledge of the meaning and it is this which itself is the delusional experience" (Jaspers: 1963, 99).That is, it is the immediate intrusive awareness of the (disturbed) meaning that constitutes the experience. And this experience of disturbed meaning has the character of reality.

It is one thing to make an error in judgment. It is something else to hold to the erroneous judgment in the face of all evidence to the contrary and against all argument. According to Jaspers, errors in normal people are the errors common to their social group (Jaspers: 1963, 105). This is why religious beliefs and common prejudices are not considered delusional. Delusion-like errors are somewhat segregated from common belief. What sets delusion proper apart from other errors, Jaspers suggests, is that the incorrigibility is not the result of partial segregation from the social group but is the result of an alteration in the personality

> *Delusion proper* is incorrigible because of an *alteration in the personality,* the nature of which we are so far unable to describe, let alone form into a concept, though we are driven to make some such presupposition. (Jaspers: 1963, 105)

Jaspers defines personality as "the *individually differing and characteristic totality of meaningful connections* in any one psychic life" (Jaspers: 1963, 428). He offers the following rationale for making the presupposition that he does, based on the 'psychologically irreducible' primary nature of the experience of delusion.

> This primary experience, moreover, is not limited to a single experience which breaks through into consciousness simply as one phenomenon among others, otherwise the patient could criticise and master it. The primary event has to be related to some radical change of personality since, otherwise, the insurmountable character of the delusion and its essentially distinctive incorrigibility would be quite incomprehensible. (Jaspers: 1963, 196)

If the meaningful connections between the patient's beliefs, emotions, desires and so forth are normal the patient would be able to correct her own errors of judgment in the same way as do normal people. The apparent incongruity between the delusional belief and the rest of their belief system suggests that the entire web of beliefs has been distorted. This gives rise to Jaspers' much contested claim that the alteration in personality makes primary delusion beyond the reach of our understanding.

4. Conclusion

Karl Jaspers' contribution to psychopathology and the study of delusions cannot be overstated. He laid the foundation for the scientific study of mental illness, and introduced a methodology that produced insights still influential today. The account I offer in the second section of this thesis is based on Jaspers' description of delusion as a disturbance in the meaning component of the immediate perceptual experience. It is implied in Jaspers' writing that he conceives of meaning as affordance (when I see a knife I see a tool for cutting), but he does not elaborate on the point. Therefore, it is unclear whether his conception of meaning is thick or thin. In my account I will argue that the meaning component of the immediate perceptual experience in the Capgras delusion is very rich. Jaspers treats the individual as a 'whole' and considers that all sciences relating to human beings are relevant to the study of mental illness. Whilst acknowledging the important contributions neurobiology and other medical sciences make to the study of mental illness, he resists the idea that mental illness can be explained by reducing it to a physical event. This is an idea I support. I acknowledge the importance of scientific research that isolates biological deficits and the contribution this makes to the explanation of delusions. But my contention is that any valid explanation of delusion will incorporate the concept of the 'whole' individual embedded in her environment. A weakness of Jaspers' account is the lack of detail regarding the mechanism that produces the false judgment that disturbs the meaning content of the immediate experience. It is my intention to address this lack by providing an account of this mechanism in the Capgras delusion. A second weakness in Jaspers' account is his insistence that the world of the delusional is beyond our understanding. I disagree with this view, as do other researchers such as Sass and Gallagher, whom I discuss in the following section.

Section Two: Following Jaspers

1. Introduction

Jasper's followers are indebted to him for his methodology and for understanding delusions fundamentally as disturbances in conscious psychic events. However, they reject his pessimism regarding our inability to understand the primary delusions predominantly associated with schizophrenia. Although patients with schizophrenia often experience monothematic delusions, the relationship between these and the polythematic delusions characteristic of schizophrenia has not been established. Researchers generally assume that explaining one will not necessarily explain the other. However, more recently, there have been attempts to extend the explanation of monothematic delusions to encompass polythematic delusions (e.g. Langdon & Coltheart: 2000; Coltheart, Langdon & McKay: 2011). My intention is to reverse this methodology by taking insights from the various phenomenological explanations of schizophrenia and incorporating them into an account of a monothematic delusion.

Phenomenological psychiatrists and philosophers, such as Sass and Gallagher, hold that concepts from the phenomenological tradition, applied with empathy and imagination, can provide a means of accessing the lived-world of schizophrenia. They each describe what they suggest are foundational aspects of the schizophrenic experience. Drawing on the works of Minkowski, Blankenburg and Kimura, Sass offers an ipseity-hyperreflexivity model (IHM) of schizophrenia. He suggests that forms of exaggerated self-consciousness cause the patient to experience aspects of self as akin to external objects, and that this gives rise to the delusional experience of schizophrenia. In keeping with the phenomenological tradition, Gallagher gives an embodied and embedded account of cognition. He offers a multiple realities theory of experience, which he suggests might provide a framework for studying and understanding

delusions. This section begins with an explication of Sass's account of the cause of schizophrenia. It is followed by a more detailed account of Gallagher's conception of cognition as embodied and embedded, and his multiple realities hypothesis. At the conclusion of each explication I will note the specific insights I incorporate into, and expand upon in my own account.

2. Louis Sass

Sass acknowledges his indebtedness to Minkowski, Kimura and Blankenburg and suggests his own theory represents a possible synthesis of their views (Sass: 2001; Sass & Parnas: 2001). According to Sass, schizophrenia is fundamentally a disturbance of ipseity.[2] Ipseity refers to the sense of existing as a subjective self-presence, that is, to "the experiential sense of being a vital and self-identical *subject* of experience of *first person perspective* on the world" (Sass: 2003, 429). The ipseity disturbance has two aspects, diminished self-affection and hyperreflexivity.

> We have described a phenomenologically oriented conception of schizophrenia as a self-or-ipseity-disorder with two main facets: decline in the sense of existing as a subject of awareness (diminished self-affection) and heightened awareness of aspects of experience that would normally remain tacit or presupposed (hyperreflexivity). (Sass & Parnas: 2001, 347)

Self-affection refers to a direct, non-inferential, non-reflective consciousness of our own occurring thoughts, perceptions, feelings, or pain, presenting in a mode that immediately reveals them as our own. "It is precisely this first personal givenness that makes the experience subjective" (Sass & Parnas: 2003, 430). To be aware of one*self* is not to apprehend a pure self that is *apart* from experience but to be acquainted with the experience from within. Because self-affection gives rise to a sense of 'myness', when there is a diminution of self-affection

[2] The suggestion that schizophrenia involves a disturbance of ipseity is drawn from the work of Kimura. Kimura sees the problem arising from a disturbance in the dual subjectivity experienced by humans, that is, between the 'me' as a purely subjective feeling of selfness and the 'I' as the publicly recognizable instance of individuality. "We are not only conscious of a self as an object of self-recognition unconsciously shared with others - as an I of "I am doing so and so" – but also that, prior to this, we are living, embodied, and feeling creatures who are subjects of a self-evident private selfhood – as the self implicit in "it is seeming to me" (Kimura: 2001, 333). Kimura suggests the problem in schizophrenia is 'simultaneous reflection' in which the self that is observed (the 'I') retains the status of a subject (the 'me'). As will be seen in the section on hyperreflexivity, Sass sees the problem as being objectification of the 'me' rather than subjectification of the 'I'.

there is also a weakening of the felt sense that experiences belong to 'me'. This strips the experience of its lived context, inviting introspective hyperreflexive awareness (Parnas & Sass: 2001). In response to the diminished sense of ownership the mind turns inwards and subjectivity becomes the object of explicit reflection.

i) Hyperreflexivity

Hyperreflexivity consists of forms of exaggerated self-consciousness in which aspects of oneself are experienced as akin to external objects. When the mind turns inwards and reflects on what would normally be tacit, there is a shift from a tacit sense of self to a *reflective awareness of* the tacit senses of self. Thus, there is a consciousness of consciousness (hyperreflexivity). Reflecting upon that which is normally tacit objectifies it, creating distance between the self that is the ground of the experience and the self that is the object of reflection. "Experientially, there is an increasing gap between the self and the flow of consciousness" (Sass & Parnas: 2003, 432), which alienates the self and diminishes the sense of body or body part ownership. If loss of ownership is exaggerated, it may manifest as delusions of alien control. Further, when patients reflect upon their own subjectivity, mental processes and inner speech are no longer permeated with a sense of selfhood.

> Phenomenological analysis suggests that such self-consciousness, a kind of introspection, is *inherently* alienating: This is because the act of observing tends to objectify, externalize or reify what is observed; when what is being observed is thinking itself, this (usually) most intimate part of the self can seem to exist at a remove from one's own intentionality. (Sass: 1999, 262-3)

With a diminished sense of ownership and the externalization of inner speech, schizophrenic patients may hear their own thoughts aloud or hear voices commenting on their current thoughts and actions (Sass: 1992). Thoughts also can be experienced as entering the mind from an external source (thought insertion), and with sufficient distancing, be experienced as existing in the public realm rather than being internal and private (thought broadcasting). Sass suggests his analysis helps clarify these first rank symptoms of schizophrenia. But the point Sass stresses regarding hyperreflexivity is that "rather than

sustaining a sense of self, certain forms of self observing may actually serve to undermine it" (Sass: 1992, 230).

ii) Body alienation

According to Sass, the sense of self and our sense of immersion in the world are inseparable. In fact we are self-aware *through* our practical absorption in the world of objects.[3]

> A tacit or subsidiary awareness of kinesthetic and proprioceptive sensations serves as the medium of prereflective selfhood, ipseity, or self-awareness, which, in turn, is the medium through which all intentional activity is realized. Any disturbance of this tacit-focal structure, or of the ipseity it implies, is likely to have subtle but broadly reverberating effects that upset the balance and shake the foundations of both self and world. (Sass & Parnas: 2003, 430)

We interact with the world from a pre-reflective sense of self that is drawn from a tacit awareness of our bodily affective states and attitudes. We live as *self-directed-towards* our world. If the body is alienated through hyperreflexivity, it diminishes the sense of self which is drawn from the body, thus disturbing the relationship between self and the world. In such case, affective experiences are deprived of an essential component of normal, targeted emotional meaning. "They cannot be part and parcel of a coherent and meaningful state of directedness towards the world … Instead they exist in a kind of free-floating space" (Sass: 2004, 135).

iii) Un-worlding

'Un-worlding' involves the schizophrenic experience of fragmented perception in which the perception of geometric unity is maintained but objects have lost their affordance. Objects no longer stand in context and may be experienced as

[3] The hypothesis that body alienation is involved in the schizophrenic condition is drawn from the work of Minkowski. For Minkowski, human subjectivity is embedded in the world with the body acting as mediator. It is our connection with the demands of reality that gives us our goals and the future oriented nature of our behaviour. A loss of connection with reality results in a loss of goals, a loss of emotional connection to the world and a loss of a basic or primal sense of vitality. Patients compensate for the diminished vital contact by overstressing the static, reflective intellectual mode of reality in their worlds, acting on this, rather than in response to the richness and movement of life (Minkowski: 1926, 1927; Minkowski & Targowla: 1923).

meaningless detail.[4] The concept of un-worlding is based on the Heideggerian understanding of 'world' as a complex unity held together by a set of instrumental meanings and relationships, that is, a self-referential system of affordance (Heidegger: 2000, 95-98). Emotional responses are stimulated by whole objects or situations and the meaning we attribute to them. If affordance is disrupted so are the emotions they elicit.

> To see a human body, or a chair, as pure three-dimensional geometry, is to forfeit the potential for reacting with lust, loathing or a yearning for peaceful repose. The un-worlding of the world is therefore synonymous with the disappearance or attenuation of many common forms of emotionality, and the normal forms of engaged temporality that emotions usually imply. (Sass: 2004, 136-137)

Sass suggests that the alteration of the cognitive/perceptual (un-worlding) and kinaesthetic/proprioceptive (body alienation) dimensions are manifestations of a more central disturbance—a disturbance of ipseity. The turning away from the world in hyperreflexivity may lead to a loss of affordance and a disturbance in the patient's relationship to her world. The exaggerated distancing in hyperreflexivity can lead to body alienation, which again disturbs the patient's relationship to her world. Thus, a disturbance in ipseity primarily affects one's in-the-moment sense of existence in the world.

iv) Subjectivization

'Subjectivization' refers to a mode of schizophrenic experience, usually occurring in the chronic stages of the illness, involving a withdrawal from the external world and social reality in favour of a preoccupation with a delusional

[4] Blankenburg (1969) suggests that schizophrenia results from a failure to develop common sense. Common sense can be defined as practical understanding. Underlying his notion of common sense is Husserl's claim that it is within the core truths of all minds that we are able to discover how all meaning is constituted (e.g.Husserl: 1981). According to this view, a disturbance in common sense places the cause of schizophrenia deep within the structure of human consciousness. Although Sass does not discuss common sense specifically, the influence of Blankenburg's work is evident in his account of un-worlding and subjectification, particularly in his suggestion that un-worlding diminishes the experience of common emotional responses to the world and the loss of normal forms of engagement that it implies. When objects lose their affordance patients lose the emotional connections common to their social environment. This reiterates the idea it is a loss of common understanding in terms of one's social environment that contribute to the schizophrenic condition.

or quasi-delusional realm of inner fantasy. The exaggerated objectification and distancing involved in hyperreflexivity means phenomena that would normally be purely abstract, mental or inner, take on a certain substantialized and externalized quality. Sass calls this 'phantom concreteness' (Sass: 1994, 14). To elicit emotions, a situation has to have some level of objective reality. According to Sass, as phantom concreteness has only a quasi-substance, the affective state produced is not a real emotion but a quasi-emotion. Because of its quasi status, the emotion evoked does not involve any real change in action readiness. This explains why schizophrenic patients so often are passive despite their belief in the sometimes terrifying content of their delusional world

It is important to the patient that the fantasy objects do not have the full ontological status of 'reality', because it is not the *content* of the world that the patient wishes to escape, but the *form* (Sass: 1994; 2004). According to Sass, patients do not confuse fantasy with reality. The maintenance of the integrity of a virtual world to which one might escape relies on an ability to distinguish between the two. Fleeing to the safety of a virtual world over which one has full control enables the elimination of all the fears, threats and demands of the real world. But this is something of a double edged sword. As creator of the world, the patient experiences a godlike omnipotence whilst at the same time experiencing a sense of vulnerability generated by the need to maintain the world they have created through a continuous flow of representing. The safety provided by the delusional world is underpinned by the fear that if the patient fails to maintain it, it may suddenly collapse. This is reflected in the classic schizophrenic fear of world-catastrophe. The distinction between the two worlds—the real and the delusional—is demonstrated by the patient's 'double bookkeeping' (Sass: 1994, 21, 43), that is, the ability to maintain a delusional belief whilst retaining a quite accurate sense of what would generally be considered to be the objective or actual circumstances.

My account of the Capgras delusion does not borrow directly from Sass's work, but his ideas influenced my early thinking about the causes of delusions. These

ideas, which appear in my account in a somewhat altered form, are mostly drawn from his description of subjectivization. His conception of delusions as a world of inner fantasy that has a quasi-reality struck me as having explanatory power. The phantom concreteness of the delusional content seemed to accommodate the fact that on the one hand delusions are sufficiently belief-like to be defined as beliefs, whilst on the other do not appear to fully function as beliefs.[5] The juxtaposing of a world of quasi-reality with the real world also seemed to offer a good explanation of the phenomenon that Sass described as double-bookkeeping. Gallagher's development of Sass's conception of delusion as an alternative reality further influenced my thinking. Gallagher's account is the subject of the next section.

3. Shaun Gallagher

Gallagher rejects the Cartesian view of the mind, which gives absolute priority to the *cogito* and which neglects the role of the body. His main argument is that mind, body and world mutually interact and influence one another to promote an organism's adaptive success. Gallagher conceives of cognition as embodied and embedded. Embodiment both limits and prescribes the types of cognitive processes available to human beings, and these cognitive processes are embedded in the environment and in social relationships. His main claim is that the way our body allows us to interact with the environment structures the way we perceive our self, our world and others. His supporting argument is based on an important distinction he makes between the *body schema* and the *body image*—two concepts which he notes are frequently confused in the literature. Gallagher also offers a theory of *multiple realities* that directly relates to the subject of delusions. Based on the Heideggerian concept of *being-in-the world*, and on the works of William James and Alfred Schutz, Gallagher suggests that just as we can enter alternative realities at certain times that take us away from our primary reality, it might be possible to enter delusional realities. And if this is true, the theory might provide a framework for the study of delusions.

[5] The doxastic nature of delusions will be discussed in more detail in a later chapter.

i) Embodied and embedded cognition

According to Gallagher, cognition, is in part, *embodied* because the structural and functional design of the body shapes the way we experience the world. For example, our upright posture and its supporting mechanisms, which free our hands for grasping and extend our visual range, constrain what counts as affordances and thereby what counts as our world. Gallagher stresses that many aspects of embodiment relevant to perception operate below the threshold of conscious perception, including automatic adjustments that regulate metabolism, heart rate, blood pressure and the like, because all play a role in the perceiver's ability to attend to or concentrate on perceiving objects.[6] "To think of perception as something that happens only to the brain is to ignore the contribution of embodiment to sensory "pre-processing"" (Gallagher: 2008, 164). As well as being embodied, cognition is also *embedded* or situated in the environment. In experience nothing exists in a vacuum. Every object and event "is always a special part, phase or aspect, of an environing experienced world – a situation" (Gallagher: 2008, 165). According to Heidegger (2000), we are already *in* the world in a way that is more basic than mere location, and being *in* the environment is one of the necessary existential characteristics of human existence. Things are not only available for our manipulation, we find ourselves already immersed in a world of objects available for manipulation, and the possibilities this offers shape our perceptions and actions.

> In the majority of our everyday dealings, we do not first encounter objects cognitively, and then decide what they are and what they can be used for. By the time we think about things, or explicitly perceive them as what they are, we have already been immersed in their pragmatic meaning. *(Gallagher: 2009b, 39).*

Perception does not involve a passive reception but an active exploration of our environment. Further, our experience is shaped by our body's capacity to move through our environment. When we want to manipulate an object in our environment, it must involve proprioceptive and kinaesthetic information about

[6] For example, if insufficient blood flows to the brain, brain cells do not receive enough oxygen and nutrients, and a person can feel lightheaded, dizzy or even faint (as in hypotension). This would impact on one's concentration and perception of objects.

our bodily situation to enable us to reach for the object to manipulate it. There is a tacit awareness of our body that always registers as an 'I can' or 'I can't'.

The concepts of embodiment and embeddedness are extended by the ideas that perception is enactive. Perception is enactive because it is essentially linked to what we do and our possibilities for action. And these are linked to what we are capable of doing, that is, perception depends on the sensorimotor skills of the perceiver.

> The idea of enactive perception puts the brain back into the body, and the body back into the world, and is therefore part of the same fabric with embodied and embedded cognition. (Gallagher: 2008, 168)

Cognition not only is enactive but also is elicited by our physical and social environment. On this view, perception is a way of coping with the environment because the first-person point of view is never a view from nowhere. It is always defined by the situation of the perceiver's body, and this situation is not merely location and posture. It is also action in pragmatic contexts and interaction with other people. Gallagher notes that according to Heidegger, when we find ourselves in the world we find ourselves already with others. Being-in-the world already includes being-with (others), and being situated already involves being situated with others (Gallagher: 2009b). Cognition involves a deeply embodied structural action but also is formed in an affective resonance generated by our surroundings and by others with whom we interact.

ii) How the Body Shapes the Mind

In his influential book *How the Body Shapes the Mind* (2005), Gallagher observes that serious confusion has arisen from a lack of definitional clarity and conflicting assumptions regarding the nature of the body schema and the body image. And "terminological and conceptual confusions persist in the more recent literature" (Gallagher: 1995, 227). His intention is to address this problem.

> I will argue that if a clear and proper distinction is made, these concepts carve up the conceptual space in a way that leads to a productive understanding of embodied consciousness. (Gallagher: 2005, 5)

The body image and the body schema refer to two different but closely related interactive systems, which he defines in the following way.

> A body image consists of a system of perceptions, attitudes, and beliefs pertaining to one's own body. In contrast, a body schema is a system of sensory-motor capacities that function without awareness or the necessity of perceptual monitoring. This conceptual distinction between body image and body schema is related respectively to the difference between having a perception of (or belief about) something and having a capacity to move (or an ability to do something). (Gallagher: 2005, 24)

Gallagher notes this distinction finds empirical ground in a double dissociation in neurology. He argues that patients with unilateral neglect, and as a consequence do not attend to one side of their body, suffer from a disruption to their perceptual body image. In contrast, deafferent patients, who receive no tactile or proprioceptive feedback from the neck down, cannot move without visually attending to their body. Their body schema is severely impaired and to move they have to replace it with a reflexive body image (Gallagher: 2005). He suggests that each concept addresses a different set of questions. The concept of body image helps to answer questions about the appearance of the body in the perceptual field whereas the concept of the body schema helps to answer questions about how the body shapes the perceptual field.

Crucial to the concept of embodiment is the idea that the body schema is outside conscious attention and perceptual awareness. In contrast to the body image, it involves a *prenoetic* performance of the body, one that helps to structure consciousness. As a result the body acquires a certain organization or style in relations with the environment. "The body *meets* stimulation and organizes it within the framework of its own pragmatic schemata" (Gallagher: 1995, 235). For example, when we walk we make all the correct adjustments necessary to remain upright and to execute locomotion without any awareness of making such adjustments, and we do not have to think how to shape our hand when we reach for a cup. As we reach for it our hand automatically takes the right shape.

> The visual observation of such objects automatically evokes the most suitable motor programme required to interact with them, and the activation of motor preparation areas in the brain form part of what it means to perceive objects. (Gallagher: 2005, 8)

According to Gallagher, although the body schema operates in a near to automatic way, it is often part of a voluntary intentional project, such as catching a ball in the context of a game. Although the actions may be explicitly willed, and at the same time governed by perception of objects in the environment, the person's complete awareness will be centred on the ball, not on the precise accomplishment of locomotion. Although sometimes we do attend specifically to some aspect or part of the body, particularly when learning a new motor skill such as dancing, most of the time our attention is directed away from the body towards the environment or towards some project we are undertaking. Thus, the intentional interests of the subject, in part, determine what reaches perceptual consciousness, which is always a selection of what is physiologically possible.

The body schema, which is basic to embodiment, is also basic to proprioceptive and ecological self-awareness (Gallagher: 1999a; 1999b). This involves the idea that information that I receive from the world includes, implicitly, information about my own self, specifically about my egocentric perspective and spatial embodiment. I have an integrated global sense of where I am spatially in relation to the immediate environment and what, in any particular situation, I am capable of doing. "In effect, ecological access provides a pre-reflective sense of self as a spatial presence and a set of capabilities" (Gallagher: 1999a, 290). He explains this further.

> Proprioceptive awareness thus provides an immediate experiential access to my pre-reflective, embodied self, even as I, as an agent, am not reflectively seeking myself, but am engaged in pragmatically and socially contextualized action. (Gallagher: 1999a, 290)

This has important implications for research methodologies for two reasons. The first is that there is a general assumption that subjects have direct access to their experience and the way in which they report on the experience is immaterial to what that experience is. But Gallagher suggests that different ways of reporting reflect different kinds of access to experience. This is indicated by tests (Marcel: 1993) which show that when subjects are asked to respond to stimuli with a blink, or with button pressing or with oral yes/no responses, there are discrepancies in the responses, with the most accurate

being from eye blink and the most inaccurate being from the oral reports. This calls into question the existence of a unitary reflexive consciousness or a unitary subject of experience responsible for reporting. It raises questions regarding the accuracy of methods of investigation that heavily rely on subjects' oral reports of their own experiences. The second involves the problem of abstraction.

> Certain traditional ways of analysing the self start off in situations that are abstract or detached from normal experience, and... the conclusions reached in such approaches are, as a result, inexact or mistaken. (Gallagher & Marcel: 1999b, 4)

Intentions that are abstract or decontextualized are those which have no purpose beyond the task itself or compliance with instructions. This is frequently the kind of intention solicited in laboratory tests during which subjects are asked to perform a simple movement which has no purpose other than compliance. This is significant because tests (Marcel: 1992) have demonstrated that differences in the nature of intentions strongly correlate with performance differences, with the worst performances being obtained in situations in which the patient was instructed to carry out a disembodied, meaningless or purely procedural action.

iii) The multiple realities hypothesis

Gallagher's *multiple realities* hypothesis is drawn from the following idea of William James (1890), which was further developed by Alfred Schutz (1974).

> The experiencing subject does not live in the one unified world of meaning that is defined objectively (in a view from nowhere), but in multiple realities, sub-universes or finite provinces of meaning. (Gallagher: 2009a, 254-2455)

James and Schutz further suggest there is a shared everyday reality or 'paramount reality' with which we normally engage but that there are multiple alternative realities that can take us away from everyday reality. These are the worlds of novels, cinema, theatre, video games and the like, and entering any of these alternate worlds requires the adoption of an alternative set of beliefs or values which are not commensurable with everyday reality. But entering an alternative reality is more than adopting a set of beliefs and values. When we become absorbed in a play, novel or film, for example, we are *in-the-world* of the play, novel or film. According to Gallagher, to be *in-the-world* of an alternative reality is more than an intellectual exercise because the change from one reality

to another requires an existential change. There is normally a clear transition as we move from one reality to another. This is particularly true when we become fully absorbed in an alternative reality for then we experience the world as having a certain presence and salience. Gallagher points to the experience of returning to the everyday world following such absorption and experiencing the everyday world as somehow "oddly unreal in relation to what I have been doing" (Gallagher: 2009a, 255).

Gallagher suggests it is quite possible that one can enter into a delusional reality just as one can enter into a dream, fictional or virtual reality. This is indicated by delusions that are more or less cut off from everyday reality, ones that are incommensurable with normal rules of reason, and ones that offer a different set of affordances. However, the alternative realities of the theatre, films, novels and games differ from delusions in that the latter are socially constructed realities designed for others and, therefore, are understandable by many people. Delusions can be quite idiosyncratic, more like dreams, and often incomprehensible to others. The incomprehensibility of delusions is well explained by the alternative reality hypothesis.

> The subject expresses the logic of the delusional reality from the perspective of that reality. From the perspective of everyday, shared reality, this may be (more or less) non-sensible. (Gallagher: 2009a, 259)

It is also true that some delusions, such as the Capgras and Cotard delusion, are not idiosyncratic in that each shares a common theme. According to Gallagher, this may be caused by neurological damage that opens up a door to a particular reality. There may be typical kinds of delusions for the same reason that there are typical scenarios developed in children's pretend games, or universal literary themes found in novels and plays.

One characteristic of delusion that Gallagher's hypothesis seems to explain is that of double bookkeeping. Double bookkeeping, first described by Sass (1994,) refers to patients' apparent ability to hold a firm belief and at the same time act as if it was not true. For example a patient who complains that members of the staff, such as doctors and nurses, are trying to poison her, happily eats anything

they put in front of her (Sass: 1994, 21). Gallagher notes that when we enter an alternative world, at some level we keep account of the fact that from the perspective of everyday reality we are playing a game (Gallagher: 2009a, 254). When at the theatre, it is this level of awareness that prevents us from leaping onto the stage to rescue a character from danger.

For Gallagher a delusion is not primarily a belief. It is not a set of false beliefs about the everyday world, or a set of odd beliefs about an alternative world. It is something experiential. His hypothesis is far more inclusive than doxastic accounts because it incorporates bottom-up and top-down influences into a broader account that includes a range of bodily, affective, social and environmental factors. In doing so, rather than offering an explanation of individual delusions, he attempts to give a more adequate characterization of delusions that would provide a framework for any such explanation.

Gallagher's multiple-realities hypothesis is fundamental to my account of the Capgras delusion, in which I argue that the 'impostor' concept providing the meaning component of the immediate perceptual experience was developed during engagement in alternative (fictional) realities. The following are further insights from Gallagher's account upon which I draw. According to Gallagher, when we enter an alternative reality there is a level at which we are aware of our actual situation. In my account I argue that there is a level at which the Capgras patient is aware that the fictional character has been inappropriately introduced into her experience of reality, and this accounts for double-bookkeeping. Gallagher points out that nothing takes place in a vacuum. Perception is not a view from 'nowhere'. It is embodied and embedded. As delusions happen in the wild, I argue that the context in which they take place is important. Where the patients' loved ones are, and what they are doing, matters. Therefore, care needs to be taken when interpreting data gathered in the restricted setting of a laboratory. Like Jaspers, Gallagher claims delusions are experiential. My account of delusion as a disturbance in the meaning component of the immediate perceptual *experience* identifies them as experiences. Further,

following Gallagher, I argue that entering an alternative reality requires an existential shift because it involves a different way of *being-in-the-world.*

The main criticism of Gallagher's multiple realities hypothesis is that it is not an explanation of delusions but a possible framework for explanations that might be developed in the future. Without such development, the value of his hypothesis remains to be demonstrated. My thesis attempts to address the criticism, at least in part, by offering a detailed explanation of the Capgras delusion that utilises Gallagher's framework.

4. Conclusion

The above writers all hold that each individual is a complex body-psyche unity embedded in the physical and social environment, and that experience is shaped by both our internal and external worlds. Delusion is seen as reflecting a fundamental disturbance in the patients' way of *being-in-the-world*, and each writer offers a different account of that disturbance. For Jaspers it is an 'alteration of personality', one which produces a 'world of new meanings'. Minkowski describes it as a 'disintegration' of personality involving a loss of vital connectedness to the world. For Kimura it is a disturbance in the relationship between 'self' and the world caused by a simultaneous consciousness of two agentive selves. Blankenburg emphasises the loss of the constraining influence of the social environment and of the 'common sense' that is founded in the 'core truths of all minds'. Sass sees it as an ipseity disorder in which hyperreflexivity leads to a loss of connection with the word, body alienation and ultimately, withdrawal from the social world. Gallagher suggests delusion might result from an existential shift from *being-in-the-world* of the everyday to *being-in-the-world* of an alternative reality.

None of the above characterisation of delusion is seen as something that can be reduced to a single biological, bottom-up causal factor. But this does not mean phenomenologists ignore biological factors.

> Quite apart from the fact that the actual treatment of human beings calls for thorough medical training, insight into the aetiology of psychic events cannot be achieved at all without some knowledge of somatic function. ... Thus psychopathology finds in neurology, internal medicine and physiology its most valuable auxiliary sciences. (Jaspers: 1963, 4)

Sass does not deny the significance of the neurobiological components of schizophrenia, nor that often we need to treat schizophrenia with pharmacological intervention (Sass: 2007, 415).

> I am opposed only to neurobiological reductionism, that is, to accounts that view neurobiological abnormalities as necessarily resulting in *mere* deficit states that are unproblematically characterizable as involving diminished rationality (rather than more subtle and complex, qualitative alterations of mental functioning); that too readily assume only one-way causation from brain to mind; or that ignore the role of culture and individual intentionality in the expression of symptoms. (Sass: 1998, 495)

Spitzer (1990) warns that reductionism fails to do justice to the complexity of human life and that while whatever can be measured should be measured, it is potentially harmful to pretend we can measure the unmeasurable. For this reason both Sass and Gallagher are working to integrate rigorous empirical laboratory testing with a phenomenological investigation of the patient's embodied experience.

The account of the Capgras delusion I present in this thesis draws heavily on the insights of the phenomenologists discussed in this section. It incorporates Jasper's contention that delusion is a disturbance of the meaning component of the immediate perceptual experience, Sass's description of delusion as a world of fantasy that produces quasi-beliefs and quasi-emotions and also, Gallagher's conception of embodied and embedded consciousness, and his multiple-realities hypothesis. The next section is devoted to the analytic tradition, various accounts of delusions it has produced, and an acknowledgement of the insights produced in this tradition that I have incorporated into my account.

CHAPTER 2: BRENDAN MAHER

1. Introduction

In the early 1970's Brendan Maher offered a new account of the aetiology of delusions that differed radically from Jasper's account. Maher makes two claims. The first is that biopathology produces an anomalous experience, which the patient finds puzzling or distressing. The second is that an explanation for the anomalous experience is sought using normal reasoning processes.

> In brief, then, a delusion is a hypothesis designed to explain unusual perceptual phenomena and developed through the operation of normal cognitive processes. (Maher: 1974, 103)

Whereas Jaspers described delusion as a disturbance in the judgments that produce meaning, Maher claims there is no impairment in the patient's thinking. Put simply, Maher sees delusion as a physical problem not a thinking problem. A second difference between Maher and Jaspers lies in their methodology. Jaspers uses phenomenology as a method of enquiry, whereas Maher is interested in employing the scientific method. However Maher also thinks understanding the patient's experience is important.

> Whatever direction we take, it will be important to keep in mind that the experimental method gives us precision of observation and measurement but cannot give us the same kind of information that we can get by trying to understand the nature of the patient's experiences. Both sorts of information are vitally necessary if our understanding of the pains of psychopathology is to advance in ways that will permit us to reduce those pains. (Maher: 2003a, 26)

The first section of this chapter begins with Maher's challenge to Jaspers. It is followed by an account of his theory and the questions it raises. The second section reviews various attempts to test his two claims.

Section One: Brendan Maher

1. Maher's Challenge to Jaspers

Maher claims his new account accommodates both Jaspers' primary and secondary delusions. Jaspers says that secondary delusions (delusion-like ideas)

are understandable in the context of the patient's experience. Arguably, the experience under consideration could be an anomalous experience that has a biopathological cause. Maher's example is a hearing impaired elderly person, who developed paranoid delusions because she thought people were whispering. According to Maher's account, the anomalous experience (with a biological cause) was the experience of people 'whispering' in the patient's presence. From this the patient (rationally) concluded others were trying to hide something from her.

According to Jaspers, in contrast to secondary delusions, primary delusions are not understandable.

> The primary delusion he defined (very problematically) as a fully-formed belief that appears in consciousness and arises without detectable antecedents. It is incomprehensible, in the sense that neither the patient not the clinician can understand how it arose. (Maher: 1999, 552)

Strictly speaking, what Jaspers described was the experience, not the belief. "In fact the primary delusion is essentially not a belief or judgement (*Urteil*) at all but rather an experience (*Erlebnis*). [...]The German is primäre *Wahnerlebnis* - primary delusional *experience*" (Walker: 1991, 98). Setting that aside, it is the incomprehensible nature of the primary delusion with which Maher strongly disagrees. He claims *all* delusions can be understood in terms of explanations of anomalous experiences.

To explain primary delusions Maher focuses on Jaspers' description of the delusional atmosphere, calling it the 'feeling of significance' (Maher: 1999). Maher's argument has three strands. The first is that the feeling of significance is a normal experience, a non-specific alarm that draws the individual's attention to a mismatch between what is expected and what is actually experienced. The second is that the feelings can be autonomously generated in the absence of any real discrepancy. The third is to speculate that the order of events is not as Jaspers suggests, that is, for Jaspers the feeling follows the delusional meaning but for Maher the delusional belief follows the feeling. As an example, Maher gives the case of a patient who looked at the marble tables in a café and suddenly became convinced that they meant the end of the world was

coming (Maher: 1999). To explain this, according to Jaspers' view, would be to say that when the patient saw the tables they were *experienced* as harbingers of the end of the world. The tables were correctly perceived as tables but there was a disturbance of their symbolic meaning. Maher suggests that the patient was already experiencing autonomously generated and unexplained feelings of significance prior to seeing the tables and the tables took on this significance (Maher: 1999). The explanation of the significance (now attached to the tables) provided the content of the delusional belief. If, as Maher suggests, the feeling of significance is generated autonomously, the repeated generation of the feeling could result in multiple delusional themes and the elaborated delusional belief webs (polythematic, elaborated delusions) characteristic of schizophrenia. The case of the woman who believed she had bees in her head, described in the next section, demonstrates how Maher's explanationist account may also be applied to delusions with a single theme and circumscribed content (monothematic delusions).

2. Maher's Arguments

i) The anomalous experience

Maher claims a series of case studies published by Southard (Southard: 1912; 1913; 1915; 1916) establishes the biopathological basis of the patient's anomalous experience (Maher: 2003a, 21). He cites two of these cases. The first describes a man who was thought to be a hypochondriac because he held beliefs such as "my stomach is full and I can't eat anything" (Maher: 1988a, 19). Later an autopsy showed he had a carcinomatous obstruction (invasive cancer) of splenic flexure of the colon. The second describes a woman who experienced distressing buzzing sounds in her head and developed the delusional belief her head was filled with bees. Her autopsy showed that she suffered from a softening of the bones of the skull, causing pressure on the brain and probably mechanical stimulation of auditory receptor areas. Southard argued that delusions constitute "crude scientific theories" about the somatic experiences that plagued the patients" (Maher: 1988a, 19).

Maher suggests that the crude scientific theories are developed in much the same way as normal belief, and serve the same purpose. "Strange events, felt to be significant, demand explanation" (Maher: 1974: 103). An anomalous experience or a perception of significance causes puzzlement. A search for an explanation is activated involving further observation. Development of an hypothesis follows. The hypothesis is rejected if new observations fail to confirm it and the procedure is repeated until an adequate match between observation and hypothesis is found. This brings a feeling of relief and reduction of dissonance. At the same time it raises the resistance to new contradictory data and lowers the threshold for recognition of confirmatory data (Maher: 1988a).

> The critical difference that causes this process to produce a delusional belief is that the experiences of deluded patients differ significantly from those of the general population. (Maher: 2003b, 18)

The belief appears to be irrational only because the anomalous data are not available to the public.

ii) The question of rationality

The main thrust of Maher's argument against the irrationality of delusional thinking is to suggest that the view is derived from comparing the patient's behaviour with that of the 'prototypical hypothetical person' who is assumed to be a rational agent (Maher: 1990).

> Real people, on the other hand, are distressingly often less intelligent and perceptive than are these ideal rational creatures. Real people's cognitions reveal an inadequate understanding of statistical principles, biases in their judgements of others, ethnocentricity in attitude, a failure to realize that religion has been replaced by science as a basis for the conduct of life, misattribution of effects to causes, distortion of reality so as to reduce dissonance, failure to recognize that altruism is selfishness, that terms such as freedom and dignity are gross self-deception, they are "mindless" when they should be "mindful" – the list goes on. (Maher: 1990, 78)

He offers an additional argument drawn from evidence of delusion-like explanations given by normal people in certain circumstances. Experiments on sensory deprivation in normal males in which subjects were deprived of sound and light (Jones: 1966) showed that minor details in the setting of a meal tray produced paranoid reactions in several subjects. From this Maher hypothesises

that paranoid delusions are a perceptual disorder (Maher: 1974). Maher claims other studies in sensory deprivation (Zubeck, Sansom & Phrsiazniuk, 1960; Zubeck *et al.,* 1962) also support the hypothesis that delusional beliefs represent the reaction of normal people, thinking normally in the presence of an unusual sensory situation that has been prolonged for some time (Maher: 1974).

Although it is generally considered that Maher's description of the thinking processes of the deluded as being rational is overstated, it represents an important shift in attitude towards the mentally ill. When the capacity for rational thought came to be regarded as the distinguishing feature of human beings, describing delusions as irrational, incomprehensible or beyond the reach of reason, carried with it the implication that the deluded were not fully human. By normalising the patient's thinking in the way that he did, Maher removed this implication.

iii) Evolution and cognition

Maher notes that cognitive operations solve problems presented by the individual's actual environment. Consequently, certain kinds of cognitive activity persist because of the benefits that result from the behaviour it generates, not because the cognition complies with the rules of logic. Illogical thinking is not selected against unless it leads to maladaptive behaviour. For example, the animal that responds to an irregularity that *may* be caused by a predator (Type 1 error) will survive better than one who considers the possibility it *may not* be a predator (Type 11 error) and waits for confirmation. Further, cognitive activity costs energy. Therefore, it is more efficient to produce optimal average solutions to recurring problems of a given class rather than the best or most rational solution to each individual case. Maher suggests people appear to be irrational because they apply generalised solutions to individual situations, and as they generalise from a much smaller sample than do scientists, the solution is not always appropriate. He points out that unlike the reasoning found in scholarly papers, practical adaptive cognitions are

developed to address the immediate reality of the individual. "They are not developed to meet the criteria of other scholars and scientists" (Maher: 1999, 548).

iv) Impaired reasoning

Maher concludes there is no *a priori* reason to assume that the patient's reasoning processes differ in some fundamental way from those of the general population (Maher: 1999, 548-9).

> There is no independent evidence of actual impairment of reasoning ability in delusional patients, apart from the inference that is made from the presence of the delusions themselves. To verify the existence of such impairment requires proof that these patients typically perform less well than other groups on tasks involving logical inference, induction, and deduction, and so forth. (Maher: 1988a, 23)

Maher claims the available data on thought disorders are insufficient to prove the case because subjects are usually tested for formal logic ability using abstract symbols. He thinks the problem with this approach is that a subject who concludes from the two premises 'all *a* is *b*' and 'all *c* is *b*' that 'all *a* is *c*' is poor at formal logic, but the subject who concludes from the premises 'all mice are animals' and 'all horses are animals' that 'all mice are horses' has failed to adequately observe her environment. We do not question people's normalcy if they are poor at formal logic but we do question it if they appear to have a deficiency in observing their environment. Importantly, this deficiency indicates a disorder of perception rather than a cognitive disability (Maher: 1974).

3. Criticisms and Responses

Maher acknowledges the following questions have been raised with regard to his account. He has attempted to answer each question, but not, I suggest, with complete success.

i) Why do most people fail to develop delusions when faced with anomalous experiences?

In reply, Maher claims that normal people do develop delusional explanations in the face of anomalous experiences. However, even if true, one might argue that

an anomalous experience would not cause a normal person to hold a belief such as 'my head is full of bees'. Maher also notes, that if a bizarre belief is held by the majority, it is not classified as a delusion.

> If a belief is held by most people, no matter how illogical it is and no matter how much it flies in the face of evidence, then it tends to be regarded as nondelusional. (Maher: 1988b, 333).

Whereas this may be true, as Jaspers notes, to hold a bizarre belief on one's own is not the same as holding one that is culturally sanctioned. That the community consider one's belief odd should be cause for reflection. Maher offers a second, more plausible answer to the question.

> One response to this is that the kinds of anomalous experience that deluded patients have appear to be much more intense and prolonged than those that occur to the population in general. Their intensity and duration is determined by the prolonged duration or frequent repetition of the pathological state creating the experience, not by brief short-term anomalies in the environment. (Maher: 1999, 566)

ii) How does the model account for delusions that occur in the absence of any anomalous experience?

Maher does not answer this question directly but suggests the problem of a substantial number of these delusions can be *partly* explained by 'coincidence'.

> The experience of surprise and puzzlement does not arise only from the failure of events to match expectations. It also seems to arise when patterns of different realms of events turn out to match each other when this was not expected. (Maher: 1988a, 27)

He notes that normal people are intrigued by these coincidences and prone to look for underlying significance where none exists. His example is the application of the Golden Ratio[7] to such objects as the facade of the Parthenon, Botticelli's Venus and the chambers of the Chambered Nautilus. He gives several clinical examples, one being the case of the man, who finding himself in front of a house numbered 11 on Armistice Day, November 11th, is so struck by the coincidence he concludes he was responsible for World War 1. He then argues that normal people are surrounded by coincidences but fail to notice them. He

[7] The problem of the Golden Ratio, solved by Pythagoras, is to determine the ratio that should exist between two lines of uneven length such that the ratio of the total length of both to the length of the longer of the two is the same as the ratio of the length of the longer to that of the shorter. For example, if you divide a line into two parts so that the longer part divided by the smaller part is also equal to the whole length divided by the longer part, you have the golden ratio.

suggests three reasons why patients do notice them. Hyper vigilance causes a close scanning of the environment, response to coincidence occurs later in the condition after patients have already developed an hypothesis to explain some experiential anomalies, and the close scanning of the environment was prompted by a feeling of significance (Maher: 1988a). Unfortunately, Maher's arguments do not address the question because they do not explain how delusions occur in the *absence* of an anomalous experience. He claims hyper vigilance is caused by a feeling of significance, which in the case of delusion is inappropriate and therefore, an anomalous experience. His second argument, that the response to significance occurs later in the condition, is also based on the presence of an anomalous experience.

iii) Why do patients develop a delusional explanation rather than a normal one?

A common criticism of Maher's model is that it fails to "describe a reason why a delusional interpretation is offered for the unusual sensory experience" (Winter & Neale: 1983, 237). But Maher claims that the patients' hypotheses are rational given their experiences and, therefore, they are not delusional explanations. According to his view, hypothesising that one's skull is full of bees is a rational explanation of the experience of buzzing sounds in one's head.

iv) Why do patients, unlike normal people, fail to reject the explanation?

Maher begins by challenging the presupposition of the question that normal people reject hypotheses in the face of evidence to the contrary.

> This criticism arises from the assumption that people free from psychopathology naturally reject implausible explanations in favour of reasonable, empirically testable ones, and/or are ready to change their beliefs when counterdata are presented. (Maher: 1988a, 25)

Maher argues this assumption is false on three grounds. The first is that scientists' reluctance to change their beliefs led Popper to assert that the conduct of science is ultimately social. The second is that deluded patients cling to their beliefs because they explain the experience better than the naturalistic ones. The third is that the media reports countless examples of weird beliefs

held by the general public (Maher: 1988a). His claim that normal people do not always reject *implausible* explanations is clearly true. However, it could be argued that normal people would reject *bizarre* explanations, such as 'my head is full of bees'. That the belief explains the experience better than the naturalistic one would not hold in all cases. The last argument is unconvincing, because religious and popular fad beliefs are held in common. They are not explanations of idiosyncratic personal experiences which are held in isolation from cultural or religious norms and often from the patients' own web of beliefs.

4. Summary and Conclusion

Maher's account proposes that delusions reflect a physical problem, not a thinking problem. According to his theory, a delusion is an hypothesis that is designed to explain an anomalous perceptual experience of biopathological origin and is developed using normal cognitive processes. It accommodates all delusions including Jaspers' primary and secondary delusions, elaborated polythematic delusion and circumscribed monothematic delusions.

Maher's new theory and emphasis on scientific methodology invigorated the debate and stimulated an increased interest in research into the causes of delusions. Most criticisms of his model relate to his claim that the patient's thinking processes are normal because it both fails to fit the empirical data and to be of sufficient explanatory power. These issues will be elaborated upon in the next section. Despite the weaknesses in Maher's model, its influence is far reaching and is evident in most analytic models presented today. It has given rise to a range of prominent theories known as 'explanationist' accounts.

Section Two: Testing Maher's Two Claims

1. Introduction

Researchers have sought to find scientific evidence to confirm or falsify Maher's two claims—that biopathy creates an anomalous experience and, secondly, that the patient explains it using normal cognitive processes. As I outline below, the first claim is supported by the frequent occurrence of delusion following brain injury and also from data provided by neuroimaging. The second claim remains controversial.

2. Biological Causes of Delusion

i) Stroke, tumour, trauma, dementing illness and disease

There is abundant evidence that delusion can follow brain injury caused by cerebral infarction (stroke), tumour, impact trauma, and dementing illness. Some of the delusions that have been reported as resulting from dementia are mirrored-self misidentification—the inability to recognize one's own reflected image (Breen, Caine & Coltheart: 2001; 2002); reduplicative paramnesia—the belief a familiar place exists in two different places at the same time (Mentis et al.: 1995); Capgras—the belief a close relative has been replaced by an impostor (Forsti, Almeida & Iacaponi: 1991) and; Fregoli—the belief that known persons are following one around in disguise, usually for reasons of persecution (Duggal: 2004). Strokes, which may or may not be associated with old age, have been linked with delusion such as Cotard—the belief one is dead or that parts of one's body do not exist (Gardner-Thorpe & Pearn: 2004), parasitosis—the patient's false belief that their skin or environment is infested with worms, insects, parasites or organisms (Nagaratnam & O'Neile: 2000; Huber et al.: 2008), Othello syndrome or delusional jealousy (Wong & Meier 1997), and paranoid delusions (Price & Mesulam: 1985). Brain tumours have also been linked with delusions (Pearl et al.: 1998; Gardner-Thorpe & Pearn: 2004). Viral disease such as acquired immune deficiency syndrome (AIDS) (Davison: 1989), viral encephalitis (Sellal et al.: 1996) and Rubella (Brown et al.: 2000) have also been linked to delusions. Traumatic brain injury (TBI), such as that suffered by accident victims and war veterans, has been linked to a wide range of delusions.

The most common delusions presenting after TBI are persecutory type delusions, although a wide variety of subtypes may also occur (Fujii & Ahmed: 2002).

ii) Neuroimaging

Research has been greatly aided by advances in brain imaging technology.

> These techniques can be used to map brain structure and function in normal human beings, and they have enlarged our knowledge of the pathophysiology of mental illnesses by demonstrating structural, metabolic, and neurochemical abnormalities in a wide range of mental disorders (Andreasen: 1988; 1381).

Magnetic resonance imaging (MRI) is used for assessing subtle changes in cerebral pathology, biospectroscopy to assess biochemical changes in the living human brain, and positron emission tomography (PET) to explore functional processes of the body, including the brain, by measuring tissue metabolic activity. According to Krishnan, the search for evidence of brain pathology in patients with psychiatric disorders has been the Holy Grail of psychiatry (Krishnan: 1990), and "delusions (or at least some of them) are eminent candidates for imaging analysis" (Ellis: 2007, 146). The literature based on neuroimaging techniques is vast. The following are merely a few examples of research into delusions involving the use of these techniques. Schizophrenia as a general condition has been a prime candidate for neuroimaging (Hechers: 2001; Cascella et al.: 2011; Tsujino et al.: 2011) as have its individual symptoms, such as paranoid delusions (Whitehead et al.: 2012), persecutory delusions (Tost et al.: 2010), delusions of reference (Menon et al.: 2011), and religious delusions (Taber & Hurley: 2007; Pun et al.: 2011). Studies of single theme delusions have also utilised neuroimaging. Examples are Fregoli (Moriyama et al.: 2007), Capgras (Lykouras et al.: 2002), Cotard (De Risio et al: 2004) mirrored-self misidentification (Villarejo et al.: 2011), thought insertion (Mullin & Spence: 2003), Othello syndrome (Graff-Radford et al.: 2011), somatoparaphrenia (Fotopolou et al.: 2011) and erotomania (Olojugba et al.: 2007). Although the neural substrate of delusional beliefs has not been clearly elucidated, delusions of misidentification, reduplicative paramnesia and delusions of a paranoid nature have been linked to lesions in the right

hemisphere, particularly posterior regions (McAllister & Ferrell: 2002; Young: 1992; O'Carroll and Prentice: 1996; Graff-Radford et al.: 2011). Any account of delusions presented today must take neuroimaging data into account.

The above research supports Maher's claim that delusion has a biopathological basis. But not all patients show evidence of organic cerebral disorder. Some researchers suggest that in these cases the disorder has yet to be discovered.

> It is possible that some of those patients identified as having no evidence of organic disease, for example many of the patients with schizophrenia, may in fact be found to have organic brain disease with more sophisticated imaging techniques, or a post mortem. (Fleminger & Burns: 1993; 26)

> Of course, one must be careful about assuming that some particular example of delusional belief does not have a neuropsychological base. (Coltheart: 2007; 1053)

But, as Coltheart points out, "many questions remain open here" (Coltheart: 2007; 1053).[8]

3. The Patient's Thinking is Normal

Maher's second claim, that the patient's thinking is normal, is contentious. The claim is counter-intuitive and it is not consistent with the bizarre content of many delusions. Researchers have sought to identify possible biases or deficits in the thinking processes of delusional patients and thus disprove Maher's claim. Maher was aware of this research and gave the following three reasons for rejecting these attempted refutations. First, if there is a basic defect in their reasoning, he claimed deluded patients would tend to arrive at false beliefs on insufficient evidence about a wide range of environmental circumstances. Second, he noted that in laboratory studies there is considerable overlap between the performances of deluded and control subjects. Third, he argued

[8] Bell, Halligan and Ellis (2008) challenged Maher's claim that an *anomalous perceptual experience* is necessary for delusions. Using the Cardiff Anomalous Scale (CAPS), they demonstrated that as a group, nonhallucinating deluded patients were not significantly different from nonclinical participants on any of the anomalous perceptual experience indices. They acknowledge that their test focused on perceptual anomalies, and did not include other distortion such as anomalous affective experience. Also, one would not necessarily expect patients with one monothematic delusion (e.g. Capgras) to register on their test as having a *general* problem with anomalous experiences.

that in the laboratory correct and incorrect inferences have no consequences, unlike the real world where reasoning is heavily influenced by the adaptive consequences of one's conclusions and the actions to which they lead (Maher: 1988a). As the laboratory environment does not adequately reflect the real world, laboratory data is not applicable to inferences made in the real world.

The salient point in Maher's account is that 'normal' thinking is not necessarily rational. Therefore, the various biases and prejudices to which delusional people are commonly prone need to fall outside the statistical range of normal to be regarded as a reasoning *deficit* as opposed to a reasoning *bias*. So it could be argued that the coupling of an anomalous perceptual experience and a *normal* reasoning bias is consistent with Maher's claim that the patient's thinking is normal. Personality variables have been investigated to determine whether or not delusional patients exhibit biases or deficits of the following kinds, and, if so, whether they could contribute to the formation and maintenance of delusions.

i) Attributional style

Attributional style is a cognitive personality variable that reflects how individuals habitually explain the events that happen in their lives. Causes may be attributed internally (to the self), or externally (others or the environment). Healthy people tend to take more credit for positive events whilst more often attributing negative events to external causes. The idea is that excessive internal or external attributions could contribute to the formation of delusions. For example, Bentall (2003) notes that negative internal (self) attribution and depression are linked. Research has demonstrated that depressed patients tend to blame themselves more when something goes wrong, and tend to credit others more for positive events than do normal people. Kaney and Bentall (1989) found that, in contrast to the depressed, paranoid patients have a much greater tendency to blame other people for negative events, and the normal tendency to attribute more positive than negative events to themselves was markedly exaggerated. Young (2000) suggests attributional bias can be applied

to the Capgras and Cotard delusions. He hypothesises that in Capgras cases a loss of affect is blamed on others (external attribution), but with the Cotard delusion the loss is blamed on the self (internal attribution). Thus, the same anomalous experience can result in different delusions depending on attributional style. However, some individuals have the opposite attributional style to the one necessary to explain the delusion, and not all individuals show evidence of a statistically deviant degree of attributional bias (Langdon & Coltheart: 2000). Further, delusions that appeal to different attributional styles cannot be fully general because there are *concurrent* cases of Capgras and Cotard in some patients (Davies *et al.*: 2001).

As promising as attributional style appears to be as a candidate for a reasoning bias factor in the causation of delusion, it has limited explanatory power. As Bentall notes "abnormal attributional style is not found in all deluded patients" (Bentall: 2003, 319). More recent studies have failed to resolve the issue.

> Although some studies have found that people with delusional thinking, delusion-prone individuals and high-risk first degree relatives of psychotic patients show an externalizing bias, others have found no difference in attributions between deluded individuals and the general population. (Ibanes-Casas & Cervilla: 20012, 86)

And despite it having been observed in patients with persecutory (paranoid) delusions, in their study McKay, Langdon and Coltheart "found no support for an association between persecutory delusions and self-serving/ externalising bias" (McKay, Langdon & Coltheart: 2005, 244). Jolley et al. concluded from their research "On their own, persecutory beliefs were not related to any particular attributional style" (Jolley et al.: 2006, 1597). The issue of attributional style remains unresolved.

ii) Conservation and observational adequacy

Stone and Young (1997) propose that a bias towards observational adequacy as against conservatism may impact on the formation of delusional beliefs. They argue that belief fixation is a conservative process, the aim being to accomplish the maximum in accommodating data at the minimum cost in overall disturbance to the previously established belief web. They suggest that in

persons experiencing delusion, this balance goes too far towards observational adequacy as against conservatism (Stone & Young: 1997). Davies and Coltheart refine this idea by describing two ways of understanding observational adequacy and two ways of interpreting Stone and Young's claim corresponding to them.

> We can thus distinguish two slightly different ways of understanding the principle of observational adequacy. On the first construal, the observational data to which belief revision should be adequate concern the external world (e.g. there is a mouse in the corner of my office) rather than my experiences. On the second construal, the data to which belief revision should be adequate are data about my experiences (e.g. I have an itch.). (Davies & Coltheart: 2000, 18)

Davies and Coltheart point out that it is not clear that Capgras patients attach too much weight to data that concern the nature of their experience. Rather, their mistake is to be too ready to adopt a particular (delusional) explanation for their experience (Davies & Coltheart: 2000).

The difficulty in assessing the likely impact this bias has on the formation of delusional beliefs relates to the content of the experience. The content of the Capgras delusion, for example, could range between extremely thin (there is something a bit odd about my relative) to very thick (that person is an impostor). If the content is merely a vague sense there is something a bit strange about the relative, it is uncertain how it would lead to the impostor hypothesis no matter how much weight was placed on the datum. If the relative is *experienced* as an impostor, Stone and Young's argument might hold.

iii) Probabilistic reasoning bias

Huq, Garety and Hemsley's (1988) describe an experiment they conducted on a probabilistic reasoning task in which deluded subjects were compared with a non-deluded psychiatric control group and a normal control group.

> Bayesian theory provides a normative model of the way in which evidence relevant to normal beliefs may be evaluated: this makes it possible to classify delusional beliefs in terms of deviations from optimal Bayesian inference. (Hemsley & Garety: 1986, 51)

The aim of the study was to determine if deluded subjects demonstrated any characteristic distortions in decision-making, deviating not only from the Bayesian norm but also from normal and psychiatric control groups (Huq,

Garety & Hemsley: 1988).[9] In the experiment they used jars labelled A and B, holding coloured beads in a ratio of 85 pink and 15 green, and the opposite ratio, respectively. Subjects were informed of the proportions and the jars removed from view. Subjects were then told that either container is equally likely to be chosen when a bead is drawn from it. Thus, the initial probability in Bayesian terms is always 50% for the hypothesis that a bead is drawn from A and 50% that it is drawn from B. A bead was drawn from one of the hidden containers and shown to the subject then replaced to maintain the ratio. This is repeated. The subject's task was to work out whether the experimenter is drawing from jar A or jar B. In 'draws to decision' the subject was free to draw as many beads as necessary to be sure about her choice. In 'probability estimates' the subject was asked to indicate after each draw the probability of one container having been chosen rather than the other. "At each stage, Bayes' theorem can be used to compute the likelihood of a given container having been selected, given the data presented" (Huq, Garety & Hemsley: 1988, 804).

> The results of this study suggest that on a neutral task deluded subjects require less evidence than normals or non-deluded psychiatric patients before reaching a decision. They also express higher levels of certainty after the first item of information is presented than either of the two control groups. (Huq, Garety & Hemsley : 1988, 809)

Probabilistic reasoning bias has become known as 'jumping to conclusions' or JTC.

Many studies using variations of this experiment have been conducted to date and the collective JTC findings are considered to be reasonably robust. It is commonly claimed that patients with delusions jump, or tend to jump to conclusions (e.g.Fine et al.: 2007; Brakoulias et al.: 2008; Averbeck et al.: 2011). This common claim can be misleading because not all patients jump to conclusions. For example, in two experiments conducted by Young and Bentall involving patients suffering from persecutory delusions and depression and normal controls, "no evidence was found of an abnormal tendency to 'jump to conclusions' in the deluded group" (Young, H. & Bentall: 1997). McKay, Langdon

[9] The Bayesian norms in these experiments were applies to conscious reasoning. The results were not intended to reflect any differences in abductive reasoning performed at the subpersonal level.

and Coltheart also found "no direct evidence that a jumping to conclusions explanatory style contributes to the formation or maintenance of persecutory delusions" (McKay, Langdon & Coltheart: 2007, 370). And they note that their study "did not ... replicate the standard association between the jumping to conclusions bias and delusionality" (McKay, Langdon & Coltheart: 2007, 371). According to Startup, Freeman & Garety "approximately one half of individuals with delusions show this data-gathering bias" (Startup, Freeman & Garety: 2008, 457). Inconsistencies in conclusions may result from the failure to distinguish between different types of delusions, the different make up of the control groups and variations in methodology. Methodology is particularly important because not all studies use 'draws to decision' and 'draws to certainty' as the criteria. Variables such as 'response to disconfirmatory evidence' (e.g. Moritz & Woodward: 2006), and 'jumping to new conclusions' (e.g. Rodlier et al.: 2011) are included in particular studies. Moritz and Woodward conclude from their study that a bias against disconfirmatory evidence "is a core feature of schizophrenia, and that this style of thinking is not confined to delusion-congruent scenarios" (Moritz & Woodward: 2006, 157). According to Rodlier et al., "the jury is still out on whether delusional thinking is associated with an inclination to 'jump to new conclusions'" (Rodlier et al. 2011, 433), but their study does show that healthy (non-psychotic) people with delusional ideation tend to change their conclusions more often than people without delusional ideation. From a meta-analysis of JTC bias studies, Fine et al. conclude that "'draws to decision' is the most reliable measure of the JTC bias" (Fine et al.: 2007, 69).

The validity of using Bayesian norms to measure probabilistic reasoning bias has been challenged. Maher points out that according to test results, deluded patients are *better* at Bayesian reasoning than are the control subjects (Maher: 2003a). Healthy people tend to be conservative and require more draws to decision than Bayesian theory requires. Therefore, patients who require *less* draws to decision fall closer to the Bayesian norm.

> Putting both studies together, it appears that patients with delusions were both superior in a reasoning task and more responsive to counterevidence than were those in normal control

> comparison groups. Perhaps a more plausible conclusion is that this kind of experiment has limited ecological validity when applied to the clinical phenomena of delusions. (Maher: 2003a, 12)

According to Chase, Hertwig and Gigerenzer, "In the past 25 years, the idea that human inference can be either defined or described by probability theory and logic has been increasingly challenged" (Chase, Hertwig & Gigerenzer: 1998, 212). They point out that the mind has evolved to tackle important adaptive problems, not to solve mathematical brain-teasers.

> In the real world, matters are more complicated than the simple content-blind norms tested in most laboratory problems assume. Here, Bayes theorem and subjective expected-utility maximization often become mathematically complex and computationally intractable. (Chase, Hertwig & Gigerenzer: 1998, 207)

From their meta-analysis of the JTC bias, Fine et al. draw the following conclusions.

> Overall, consideration of our findings and the possible role of JTC bias could take in delusion, suggests that JTC bias is not relevant to the formulation of delusional hypotheses. ... Nor is the JTC bias likely to play a role either in an explanation of why patients take these implausible hypotheses seriously or in the explanation of how delusional patients identify putative evidence for the hypotheses. Finally, the JTC bias seems to have no role to play in the maintenance of delusion in the face of disconfirmatory evidence. (Fine et al.: 2007, 74)

The above conclusion is in keeping with Maher's suggestion that a more plausible conclusion from research results is that JTC has little ecological validity when applied to the clinical phenomena of delusions. Research into JTC bias continues today.

4. Summary and Conclusion

Maher's two claims—that biopathy produces an anomalous experience in deluded subjects, and that they seek an explanation using normal reasoning processes—have been the subject of ongoing research in the analytic approach. Evidence from cases of brain injury and from neuroimaging data support his first claim, although cases in which there is no evidence of brain impairment suggest that not all delusions fit his model. The support for his second claim is less convincing. The results from tests investigating a range of thinking biases have proved inconclusive. However, "Overall it seems then that Maher is right in saying a failure of rational and/or logical thinking is not characteristic of deluded patients" (Langdon & Coltheart: 2000, 201). The quest to resolve the

problems in Maher's 'explanationist' account has dominated analytic literature in the past few decades. But it is widely accepted that a single factor (the anomalous experience) is not sufficient because there are cases in which the anomalous experience is present but not the delusion. Also, it fails to explain why a patient whose reasoning is normal does not reject the bizarre belief once formed. Consequently, most, but not all, recent accounts include at least two causal factors in the aetiology of delusions. The account I offer of the Capgras delusion incorporates Maher's claim that the primary aetiology in delusions is a biological anomaly. In my account the anomaly is impairment of face encoding. In keeping with more current thinking, I support the idea that a second factor is necessary. In the next chapter I will describe two two-factor accounts before describing two more recent one-factor accounts.

CHAPTER 3: TWO-FACTOR ACCOUNTS

1. Introduction

The strength of Maher's account is its suggestion that delusion might be an explanation of an anomalous experience with a biological origin. But a weakness of his account is that it fails to adequately explain why delusional patients seem unable to revise their bizarre beliefs in the face of clear and incontrovertible evidence to the contrary. Also, there is clear evidence that some people have the same neuropsychological impairment, and, therefore, very probably the same anomalous experience, but do not develop a delusion. For these two reasons those supporting an 'explanationist' account usually deem the first factor necessary but not sufficient for the formation and maintenance of a delusion. A second factor is needed. Evidence and arguments for the two-factor theory have been drawn predominantly from the study of monothematic delusions and it is, therefore, a two-factor theory of *monothematic* delusions. There have been some attempts to extend the theory to encompass polythematic delusions but research in this area is in its early stages.

Two-factor theories have been prominent in the analytic literature for the last decade, but during that time some researchers have offered a one-factor account that also addresses the weaknesses in Maher's account. In this chapter I will describe two current two-factor accounts, one offered by Coltheart, the other by Davies and Davies. In the following chapter I will describe two one-factor accounts, one offered by Gerrans, the other by Fletcher and colleagues.

2. Two-factor accounts

Two-factor accounts accommodate both biases and deficits. *Factor* is a general term referring to any perceived cause of the formation and maintenance of a delusion. A *deficit* refers to a factor thought to be an impairment to, or loss of, some aspect of neuropsychological function. A *bias* is a tendency to use information in one way rather than in another. One delusion purported to be

caused by a *deficit* is the Capgras delusion, which is thought to result from impairment of the autonomic response to familiar faces. A delusion thought to be caused by a *bias* is delusional jealousy, as biased interpretation of selective data over time is sufficient to explain the content of the delusion.

The two-factor cognitive neuropsychological account was introduced by Langdon and Coltheart in 2000 and conceived of as a two-deficit theory that could be applied to all delusions.[10] According to the theory, the first deficit accounts for the content of the delusion. It is different for each type of delusion. The second deficit accounts for its persistence and is common to all delusions. It is hypothesised to be impairment to a belief evaluation system. The model is applied to monothematic delusions, but Langdon and Coltheart suggest that "polythematic delusions may arise when a failure of normal belief evaluation is coupled with a multitude of different perceptual aberrations" (Langdon & Coltheart: 2000, 210). To demonstrate the veracity of their theory, evidence supporting the existence of both deficits needs to be provided. To date, whilst plausible explanations have been offered for several monothematic delusions, there are still many delusions (e.g. erotomania and delusional jealousy) where a plausible folk psychological story can be told, but for which a plausible neuropsychological account is still to be found. In most cases for which plausible neuropsychological accounts have been given, there are either no data, or insufficient data, to prove the case. Currently, Langdon views delusions as taking place in a two-factor *framework* in which any two combinations of deficits and biases may be used to explain a specific delusion.[11] Coltheart continues to pursue a two-deficit model although he concedes that "in its most general formulation, the two-factor theory of delusional belief does not require the two factors to be due to neuropsychological abnormalities" (Coltheart: 2010, 18). In the following section Coltheart's two-deficit theory will be discussed.

[10] After its introduction, the two-factor account was developed further in conjunction with Martin Davies, Ryan McKay and Nora Breen.

[11] Personal communication: Feb 25th 2012

3. Max Coltheart's two deficits theory

Coltheart says that in order to explain any delusion we have to answer two questions.

> The first question is always: *where did the delusion come from?* –that is, what is responsible for the *content* of the delusional belief? The second question is always: *why does the patient not reject the belief?* (given that these beliefs are frequently bizarre, that there is good evidence that the belief is false, and that everyone around the patient will be denying that the belief is true) –that is, what is responsible for the *persistence* of the belief? (Coltheart: 2007, 1044)

This suggests two factors might be necessary for all delusions, one to answer the first question and the other to answer the second question. Coltheart conceives of these factors as deficits.

> *Deficit A:* The patient has a neuropsychological deficit of a kind that could plausibly be related to the content of the patient's particular delusion – that is, a deficit that could plausibly be viewed as having prompted the initial thought that turned into a delusional belief.
> *Deficit B:* The patient has right-hemisphere damage (i.e., damage to a putative belief evaluation system located in that hemisphere). (Coltheart: 2007, 1047)

Coltheart's two-deficit hypothesis centres on a group of monothematic delusions for which possible neuropsychological causes have been hypothesised or identified. The following account refers specifically to this restricted group of delusions. Coltheart makes the point that within the schizophrenic condition delusions are usually polythematic, but patients may also exhibit a monothematic delusion such as the Capgras delusion. Where this occurs, the monothematic delusion is independent of the schizophrenic condition and the same two-deficit account applies (Coltheart, Langdon & McKay: 2007). For Coltheart's account to be plausible he needs to demonstrate two things. The first is that he needs to show that for each delusion there is neuropsychological impairment of a kind that is likely to give rise to the content of the delusion. The second is that he needs to show that there is neurological damage to an area of the brain which plausibly, could be the location of a belief evaluation system.

i) Factor one (deficit A)

The poster child for the two-deficit theory is the Capgras delusion.[12] Testing has demonstrated that Capgras patients as a group record reduced skin

[12] For an expanded account of the Capgras delusion see Chapter 6.

conductance responses to familiar faces.[13] This indicates the presence of some kind of neurological anomaly that could plausibly function as the first factor. In addition, most brains scans of Capgras patients show lesions in the right frontal lobes, thus providing a possible location for the second factor. In Maher's account and Coltheart's earlier account (Coltheart, Langdon & Breen: 1997; Langdon & Coltheart: 2000; Davies & Coltheart: 2000) the content of the delusional belief is produced by attempting to explain an anomalous *experience.* For the Capgras patient, it was originally thought that the anomalous experience arose because the relative *looked* familiar but did not *feel* familiar. However, this theory did not hold because people are not conscious of the activity of their autonomic nervous system. Therefore, in his more recent account, Coltheart locates the explanation of the anomaly at the subpersonal level rather than at the level of consciousness. "In many forms of delusion, something abnormal occurs of which a person is not conscious" (Coltheart, Menzies & Sutton: 2010, 264-5). But he stresses that he does not want to suggest that *every* type of delusion is triggered by an unconscious abnormality. For delusions of reference and persecutory delusions he says the initial abnormality which triggers the inference process may be a conscious experience.

In support of his theory Coltheart offers six additional monothematic delusions for which he describes a putative impairment. He claims they all demonstrate a plausible connection between anomalous data and the content of the specific belief. The following list is taken from *Abductive inferences and delusion belief (*Coltheart, Menzies & Sutton: 2010, *Table 1*, 267-268).

- **Fregoli:** Autonomic system is over-responsive to faces, thus, even the faces of strangers produce autonomic responses, leading to the belief *'people I know are following me around but I don't recognize them because they are in disguise'*.

[13] The familiar faces are primarily those of close relatives. The reduced response is not recorded for famous faces. SCR is thought to reflect affective response.

- **Cotard:** Autonomic system is under-responsive to all stimuli, thus, there is no emotional response to one's environment, leading to the belief *'I am dead'*.
- **Mirrored-self misidentification (with mirror agnosia):** Mirrors are treated as windows, leading to the belief *'that can't be me'*.
- **Mirrored-self misidentification (without mirror agnosia):** Impaired face processing causes a mismatch between the reflection and the representation of the patient's face stored in long-term memory, leading to the belief *'this isn't me'*.
- **Alien control:** Failure of computation of sensory feedback from movement, thus, voluntary movement is no longer sensed as voluntary, leading to the belief *'other people can cause my arm to move'*.
- **Somatoparaphrenia:** Damage to motor area of the brain controlling arm, thus, left arm is paralysed and the patient can't move it, leading to the belief *'the arm is not mine, it is [some specified other person's arm]'*.

Coltheart acknowledges that the two-factor accounts of Fregoli and Cotard delusion are 'largely speculative' although 'perfectly testable', and that somatoparaphrenia is complicated by the fact that it is not the only delusion associated with paralysis (Coltheart, Langdon & McKay: 2011).

Having argued for the plausibility of a monothematic delusion being an explanation of anomalous subpersonal data, Coltheart (Coltheart: 2007; Coltheart, Menzies & Sutton: 2010) expands his account by hypothesising how the explanation is derived. The account is based on a model of cognition which theorises that we continually generate predictions concerning what will happens to us next on the basis of what we currently know about the world.

> These predictions are normally confirmed. But occasionally they are not, and this calls for an inspection of the database of beliefs about the world, which we use to generate the predictions. The processes of prediction, and the detection of failures of predictions, are all automatic and unconscious. (Coltheart: 2005, 155)

When a prediction is falsified, a subpersonal system generates hypotheses which would account for the unpredicted event. In the case of Capgras, the prediction would be that a familiar face will be accompanied by an autonomic

response to familiarity. If the expected autonomic response does not occur, then the prediction fails and the system generates suitable hypotheses which, if true, would explain the actual state of affairs. The hypotheses are then compared on the basis of probability given the relevant factors, and the hypothesis that best fits the observed data is selected for elevation to the status of belief.

Coltheart speculates that the hypothesis is selected using abductive inference. Abductive inference (also known as inference to the best explanation) is a form of non-deductive reasoning that selects the best hypothesis from a range of hypotheses that explain the observed data. Abductive inference differs from other types of inference in that the conclusion introduces new concepts that are not present in the observational evidence. It is divided into two stages. The first stage starts with an observation (B) from which an inference is generated in the following way.

- B has been observed
- If A had been the case, B would have been observed
- Therefore, A was the case

If we apply this to the Capgras delusions we arrive at the following.

- A lack of predicted autonomic response to a familiar face has been observed (B)
- If the familiar person is in fact a stranger then B would be observed
- Therefore, this person is a stranger

Stage two consists of considering alternative hypotheses and assessing which best explains the observed data. This is where the Bayesian model of belief system revision becomes important to Coltheart's account, because it accommodates degrees of probability by giving weightings to an hypothesis' explanatory power. Hypotheses are assessed, not in terms of data confirmation, but in terms of which hypothesis is more *probable* given the data. In the Bayesian account the rival hypotheses are assessed, in part, by comparing the prior probability of the hypotheses being true. As an example of a hypothesis with a low prior probability, Coltheart gives the example of the hypothesis that a miracle has occurred. Although this hypothesis may best fit the fact that a

witness has reported seeing a miracle, the prior probability of a miracle that violates the laws of nature having occurred is very low (Coltheart, Menzies & Sutton: 2010).

Coltheart's challenge is to apply the above to the Capgras delusion in a way that demonstrates how the patient arrives at the belief his wife is a stranger, given the low prior probability that this would be so. Given the two hypotheses, 'that person is my wife' and 'that person is a stranger', one would assume the first would be judged more probable in terms of prior probabilities. Coltheart meets this challenge by pointing out that in Bayesian terms it is not the prior probability but the posterior probability that determines the acceptability of an hypothesis. To determine the acceptability of an hypothesis, one needs to take into account its prior probability and the likelihood of the observed data given the hypothesis. Bayes' theorem provides a formula for making such a calculation. According to his theorem, the posterior probability of an hypothesis can still be high even when its prior possibility is low, providing the likelihood function is sufficiently high. Using Bayes theorem and speculative weightings, Coltheart calculates the comparative probabilities of the two Capgras hypotheses. In this way he demonstrates that the posterior probability of the impostor hypothesis may be a hundred times greater than the posterior probability of the wife hypothesis.

> The general point here is that if the stranger hypothesis explains the observed data much better than the wife hypothesis, the fact that the stranger hypothesis has a lower prior probability than the wife hypothesis can be offset in the calculation of posterior probabilities. And indeed it seems reasonable to suppose that this is precisely the situation with the subject suffering from Capgras delusion. His delusional belief provides a much more convincing explanation of his highly unusual observations than the non-delusional belief. So if the subject with Capgras delusion reasons in this way, he has committed no mistake of rationality. (Coltheart, Menzies & Sutton: 2010, 278)

Once a hypothesis has been selected in this way it is accepted as a belief and conveyed to consciousness. This explanation takes us from anomalous data to the formation of the belief. But we are still left with the questions of why, once the patient is aware of the belief, it is not rejected, and why some people have the neuropsychological impairment but not the delusion.

ii) Factor two (deficit B)

The fact that patients do not reject bizarre beliefs leads Coltheart to hypothesise that the second deficit is damage to a belief evaluation system, the purpose of which is to update the belief system on the basis of evidence relating to any particular belief. Whereas the first deficit, which gives rise to the content of the delusion, operates at the subpersonal level, the second deficit impacts on data obtained at the conscious level. For example, the following may be the case for the Capgras patient.

> The subject might learn that trusted friends and family believe the person is his wife, that this person wears a wedding ring that has his wife's initials engraved in it, that his wife has not been reported missing, and so on. (Coltheart, Menzies & Sutton: 2010, 279)

If these further data are processed using the comparative form of Bayes' theorem, the fact that the wife hypothesis explains the new data much better than the stranger hypothesis outweighs the success of the stranger hypothesis which resulted from the initial calculation performed at the subpersonal level. The patient should revise his belief accordingly. The fact that he fails to do so leads Coltheart to conclude that the patient is ignoring or disregarding any new evidence that cannot be explained by the stranger hypothesis. "So it seems as if the new information does not even enter the deluded subject's belief system as data that needs to be explained" (Coltheart, Menzies & Sutton: 2010, 280).

According to Coltheart, this account both explains why the Capgras patient maintains the delusion and why the ventromedial patient, who has the same disconnection between the face recognition system and affective response system, does not become delusional. Unlike the Capgras patient, the ventromedial patient re-evaluates the belief in the face of new information. But to make this argument work it is necessary to assume that the ventromedial patient has the *same initial belief* as the Capgras patient, that is, that his wife is a stranger. This is indeed what Coltheart suggests, but notes "we know of no studies that could have provided evidence in support of or conflicting with this conjecture." (Coltheart, Menzies & Sutton: 2010**:** 281).

Coltheart's hypothesis that the belief evaluation system is located in the right frontal lobe is based on brain imaging data that show many patients with monothematic delusions have damage to this area. But for Coltheart's account to be convincing, evidence of right hemisphere damage needs to be found in all, or nearly all, cases of a specific delusion, and this is far from being the case. In a review of sixty-nine case reports of Capgras delusion for which brain imaging results are available, the results for twenty-seven patients were normal (Dulai & Kelly: 2009). Having said that, the forty-two cases in which damage *was* recorded, are statistically significant.

iii) Nonneuropsychological delusions

The two-deficit theory was initially intended to apply to all delusions, but the lack of data indicating neuropsychological impairment in polythematic delusions such as those of reference, grandeur and persecution, caused them to be exempted. Monothematic delusions, such as erotomania, for which a plausible motivational story can be told, but not a neuropsychological one, were also exempted. More recently, Coltheart (2010) has questioned these exemptions. Citing a review of 29 cases described by Anderson, Camp & Filley (1998), he claims "there are cases where erotomania arises as a consequence of neuropsychological damage" (Coltheart: 2010, 24). Although in a minority of the cases cited the delusion did follow shortly after the onset of neuropsychological damage, a causal connection has not been established. In most cases the delusion occurs much later than the diagnosis of the medical condition, making a causal relationship difficult to establish. In addition, the authors of the review note there was "no predominant site of cerebral pathology" (Anderson, Camp & Filley: 1998, 336) and conclude "the relationship between neurological and medical conditions remains uncertain" (Anderson, Camp & Filley: 1998, 335).

Coltheart challenges the non-neuropsychological status of delusions of reference by citing Kapur's (2003) proposal that all delusions have a neurochemical basis. Kapur suggests a dysregulation in dopamine produces

inappropriate salience and that delusions are the result of patients trying to make sense of aberrantly salient experiences. More specifically, he suggests altered salience of environmental stimulus, arising from a hypothesised disturbed firing in the mesolimbic dopamine system, causes significance to be attributed to non-significant objects and events. If true, delusions of reference could be explained in two-deficit terms.[14] Various theories of dopamine dysregulation have been put forward. However, Kendler and Schaffner (2011) conclude from their analysis of thirty years of research into the dopamine hypothesis of schizophrenia (DHS), "As a scientific theory, the DHS has to date performed relatively poorly and has, from a perspective of leading models of science, important deficiencies" (Kendler & Schaffner: 2011, 59). One deficiency they note is that the theory has become so broad that it is no longer falsifiable. Having said that, they acknowledge dopamine theories have heuristic value in that they stimulate research. It gives Coltheart a plausible story to tell about delusions of reference, but like most of the two-deficit stories being told, it has yet to be confirmed by scientific data.

Coltheart gives alien abduction as an example of a nonneuropsychological delusion. He speculates that the anomalous experience is hypnopompic hallucinations and sleep paralysis that many healthy people experience upon waking.[15] He argues that these are the kind of sensations that might be expected if someone was actually being abducted, and therefore "it is not implausible to attribute the generation of alien abduction to abductive inference concerning these sensations" (Coltheart, Langdon & McKay: 2011, 291). He argues further, that those that develop the delusion are those who hold a wide range of New Age beliefs, therefore, impairment to a belief evaluation system is not necessary. This is an interesting example because the delusion does not fit

[14] Delusions of reference are one of the first ranked symptoms of schizophrenia.

[15] Hynopompic hallucinations are vivid hallucinations, often accompanied by an acute sense of danger, that occur during sleep paralysis. Sleep paralysis occurs naturally during rapid eye movement (REM) sleep. Sometimes people become aware before the REM cycle is complete, at which time their awareness of paralysis may be accompanied by hallucinations. Hallucination of strange figures and objects are common.

either a two-deficit or two-factor model, as only one non-deficit factor is needed to produce the delusion.

iv) Discussion

The empirical data linking brain injury and delusion provide some support for Coltheart's deficit theory. But the way he fleshes out the idea is not without problems. In the absence of relevant empirical data, Coltheart provides 'plausible stories' for some delusions, which link the anomalous data with the content of the delusional belief. However, for the majority of monothematic delusions a plausible story has yet to be told. Examples are: the delusion of subjective doubles (belief in the existence of a double of oneself living a separate life), intermetamorphasis (belief one has switched identities with someone else, or that others believe this to be so), thought insertion (alien thoughts are being inserted by another agent), thought influencing (thoughts are controlled by an external agent), thought withdrawal (thoughts are being withdrawn from one's head), thought broadcasting (everyone can hear one's thoughts), the Othello syndrome (delusion of unfaithfulness in a faithful partner), reverse-Othello (delusion of faithfulness in an unfaithful partner, or one who has left), erotomania (one is loved from afar by another, usually of higher social status), and anosognosia (denial of hemiplegia). These delusions resist attempts to identify the underlying abnormal data, and no agreement on a plausible hypothesis has yet been reached. Coltheart does provide a plausible story for the Capgras delusion and this will be discussed in detail in the second part of my thesis.

Coltheart's use of Bayes theorem to explain the formation of the delusional belief is not always convincing. According to the theory, the hypothesis is selected for belief status on the grounds that, if true, it would explain the actual state of affairs. He gives the following account of the Cotard delusion.

> It has been speculated by Ramachandran & Blakeslee (1998) that what causes this delusion is that the autonomic nervous system is severely underresponsive to any form of stimulus, so that strong autonomic responses are never evoked: That is the datum D. Step 2: If I were dead, that is proposition P which if true would explain D: so it is a legitimate abductive inference. (Coltheart, Langdon & McKay: 2011, 286)

But the belief 'I am dead' does not fit the state of affairs (loss of strong autonomic response) as death, presumably, involves a loss of *all* autonomic responses, not merely the loss of strong ones. Hypothesis such as 'I am depressed', 'I have lost all interest in life' or 'I am ill' seem far better choices. There are other delusions, such as anosognosia, that Coltheart's proposal cannot accommodate. Anosognosia is a condition in which the patient appears to be unaware of a disability, such as blindness or paralysis. In cases of anosognosia it is not at all apparent how the paralysis of one's arm is explained by the belief that one's arm is not paralysed.

4. Anne Davies and Martin Davies – parametric variations

Davies and Davies suggest that the heterogeneity in explanations of delusions can be conceived as parametric variation within the two-factor framework. They aim to extend the two-factor framework in such a way as to offer the prospect of reasonable coverage without so overgeneralising it that it encompasses all false beliefs. In light of their account, they note the following.

> The scope of the two-factor framework might gradually be extended from neuropsychological cases of monothematic delusion to include cases of delusion without apparent brain injury and, ultimately, the floridly elaborated delusional system of some individuals with schizophrenia. (Davies & Davies: 2009, 290)

They begin their account with the idea that "the two-factor framework is also a three-stage framework" (Davies & Davies: 2009, 290), thus making explicit that which is implicit in the Coltheart account. They note "there is a processing stage leading from the unusual experience or neuropsychological deficit to the initial adoption of the false belief" (Davies *et al*.: 2009, 193).The stages are; (Stage 1.) a deficit or an anomalous/ambiguous experience; (Stage 2.) belief formation by acceptance of the experience as veridical or selection of the most appropriate hypothesis; (Stage 3.) impaired belief evaluation. They argue that impairment to a belief evaluation system is, in fact, impairment of working memory or executive function with a neural basis in damage to the right frontal region of the brain.

i) The first factor

Davies *et al.* (2009) point out that "neither an unusual experience nor a neuropsychological deficit provides a complete answer to the question of where the delusion came from" (Davies *et al.* 2009, 193). There is a processing stage that leads, by personal-level processes of explanation or endorsement, from the deficit or the anomalous experience to the belief. They hypothesise this processing stage is subject to parametric variations.

As an anomalous experience may be either *explained* or *endorsed* (taken as veridical), they suggest it is one locus of parametric variation. When the experience *fully encodes* the content of the delusional experience, the content may be endorsed erroneously for one of two reasons. It may be endorsed because of the failure of the prepotent doxastic response (Bayne & Pacherie: 2004), or because a bias towards observational adequacy against conservatism may cause one to endorse what is seen in preference to prior beliefs (Stone & Young: 1997). When the representational content of the experience is *less specific* the process from experience to belief must involve a substantive explanatory process of hypothesis generation and confirmation. This provides the locus for further parametric variation because the formation of the hypothesis that is confirmed as a belief might be *normal*, *biased,* or *impaired*. Thus, the hypothesis, which they take to be selected using Bayes theorem, may be distorted by various reasoning biases such as attributional bias or jumping to conclusions (JTC). An additional variation that applies to both the explanationist and endorsement account is that the experience may not be abnormal but merely ambiguous. An ambiguous experience may lead to a *flawed explanation* (explanationist account) or to *misinterpretation* of the experience (endorsement account). A further variation they consider is that the route from the anomalous experience to the belief may lie mainly at the subpersonal level or mainly at the personal level. Davies and Davies suggest the explanation/endorsement parameter is likely to be a matter of degree. There seems no reason why this sliding scale might not be true of other parametric variations. Although the

explanatory variation can give rise to further parametric variations, a second variation, can only be present to the degree that the first is present (Davies & Davies: 2009).

ii) The second factor

The purpose of the second factor, impairment to a belief evaluation system, is to explain why the delusional belief is not rejected. Davies and Davies offer the following account of the impairment. A belief is formed either by accepting an experience as veridical, or by forming and evaluating hypotheses and accepting the one that offers the most adequate explanation of the experience.

> In order to consider and evaluate alternate explanatory hypotheses, the patient first needs to take a step back. The patient must inhibit the prepotent doxastic response of treating a perceptual experience as veridical and, instead, treat the experience as standing in need of explanation. (Davies & Davies: 2009, 302)

According to Davies and Davies, the second factor involves two cognitive processes—executive function and working memory. The inhibition of the prepotent doxastic response is an executive process, and evaluation of competing hypotheses utilises working memory because it requires the manipulation of information. This gives rise to their proposal that the second factor (Stage 3.) is impairment of working memory or of executive function, with a neural basis in damage to the right frontal region of the brain.

To cash out this idea they turn to dual process accounts of reasoning, judgment and decision making, sometimes referred to as System 1 and System 2 processes. System 1 (heuristic) is rapid, preconscious and computationally powerful. System 2 (analytic) is slow, sequential and effortful. System 2 processes permit hypothetical thinking that cannot be achieved by System 1. Davies and Davies hypothesise that the imperative of observational and explanatory adequacy and conservatism belong to System 1 and the assessment of competing hypotheses belongs with System 2. "A central idea in dual-process accounts is that the two kinds of processes can come into conflict or competition" (Davies & Davies: 2009, 305). Tests on conflict tasks (De Neys: 2006), in which subjects had to perform a secondary task that burdened

executive resources, demonstrated that when these two processes come into conflict, greater working-memory capacity results in better performance on conflict items, but executive load has a negative impact on performance. This supports the idea that when the imperatives of observational and explanatory adequacy and conservatism (System 1.) conflict with the normative requirements of belief evaluation (System 2.), performance will be negatively impacted by limitations in executive working-memory resources.

> We may also suggest that *erroneous* belief evaluation – maintaining a false belief- in cases where the normative requirements of belief evaluation *conflict* with the imperatives of observational adequacy, explanatory adequacy, or conservatism is *caused by limitations* in executive working-memory resources (that is, working memory and executive function). (Davies & Davies: 2009, 307)

To ground this idea in neurology, Davies and Davies cite the investigation of the neural basis of performance in subjects participating in a belief-bias experiment (Goel & Dolan: 2003). This study shows that when there is a conflict between validity and belief based response, subjects who yield to prior beliefs and give a logically incorrect response show activation of ventro-medial prefrontal cortex whilst subjects who give a logically correct response show activation of the right inferior prefrontal cortex.

> One possibility is that this neural activation corresponds to controlled or executive attention that is required to facilitate the performance of the logical task of assessing the validity of the argument in the face of distraction from prior beliefs about the conclusion. (Davies & Davies: 2009, 308)

They suggest that the finding of right prefrontal cortex activation, when subjects give logically correct responses under conditions of conflict, seems to be broadly consistent with the suggestion that belief evaluation has a neural basis in the right frontal region of the brain. Consequently, right hemisphere damage is the neural correlate of the impairment to belief evaluation.

Davies and Davies apply their theory to delusions by giving the following account of anosognosia—in this case denial of hemiplegia—in which a conflict arises between the reality of the patient's situation and the patient's long-held beliefs. Normally the patient's awareness of paralysis would override their belief web, leading to belief revision. In their account, the first factor in the formation of the delusion is not the paralysis itself, but an impairment of the

patient's immediate experience of the paralysis. They give as examples of causes of this impairment: memory impairment specific to information about the movement and position of parts of one's body; information provided in bodily experiences are not consolidated into more lasting representations and; failure to integrate input signals related to self-awareness and one's beliefs about the functioning of contralateral body parts. In addition, possible neuropsychological deficits that could cause unawareness of motoric failure are unilateral neglect and impairment to the intentional-preparatory systems involved in motor control. They suggest that in such cases the conflict would be severe.

> A patient with motor impairment but without impairment of sensation or attention, intentional-preparatory systems or comparators system, memory or consolidation, would very probably recognize his or her paralysis, and would do so relatively immediately, without depending heavily on working memory or executive processing. However, recognition of paralysis is more demanding when, as a result of one or more of these impairments, it is not phenomenologically immediate. (Davies & Davies: 2009, 312)

Davies and Davies predict that in these cases, limitations of executive working-memory resources may well have the consequence that prior beliefs that are now false may be retained rather than rejected. They cite two studies of unilateral neglect (Amiola: 1999; Maguire & Ogden: 2002), regarded as a possible first factor in anosognosia, in which all patients showed impairments in at least two of the following; visual or verbal memory, sustained attention, working memory and executive function. They conclude

> These findings are broadly consistent with the proposal that the second factor in the explanation of delusions, including anosognosia for motor impairments, is an impairment of working memory or executive function with a neural basis in damage to the right frontal region of the brain. (Davies & Davies: 2009, 314)

iii) Discussion

The Davies and Davies account offers a new interpretation of the formation of the delusional belief (Stage 2.), given an anomalous or ambiguous experience (Stage 1.). They also give an account of the patient's apparent failure to evaluate or revise the delusional belief (Stage 3.). They describe the delusional belief formation as a complex combination of parametric variations and degrees of variation that differ from delusion to delusion, and from case to case. Although their account has broad explanatory power, owing to the multiple combinations

of parametric values available, this same multiplicity makes it unclear how it could be tested. As, in each individual case, it is unknown if the anomalous experience is *explained* or *endorsed*, if the representational content is *specific* or *less specific*, or whether the belief is *normal, biased,* or *impaired*, it is impossible to know precisely what it is that we should be testing. Any test we might devise that is specific enough to generate informative data would not be general enough to apply to all cases and we would be unable to determine to which patients the test would apply. In cases such as Capgras where the precise content of the anomalous experience is unknown, it is impossible to determine to what degree the belief is an endorsement of content and to what degree it is an explanation. If there is, in fact, an explanation component, then, as the explanation component is subject to further parametric variations, it is impossible to determine how significant these are as contributing factors. Consequently, the proposed account does not have implications for treatment that addresses the cause or causes at this stage. This is unfortunate because their account is a welcome attempt to address the complexity of delusions. Their proposed cause of the failure of belief evaluation offers something more substantial. The hypothesised impairment of executive function and working memory could give rise to predictions that are testable. For example, as they propose that the impairment has a neural basis in damage to the right frontal region of the brain, it could be predicted that testing working memory under conditions of executive load would produce atypical activity in this area of the brain. Presenting the two-factor framework in three stages brings increased clarity to the Davies' account. I use this strategy in my own account, although I do not present it as a feature of the account. Nevertheless, each stage is clearly delineated from the next, making my account a two-factor multiple-stage account.

5. Conclusion

The two-factor account addresses two weaknesses in Maher's account—his failure to give a plausible explanation for the persistence of delusions in the face of contradictory evidence, and his failure to explain why some people have the

same anomalous experience but not the delusion. The two-*deficit* account is elegant, and its simplicity makes it appealing. Unfortunately, the empirical data available at present are more suggestive than convincing, and I think attempting to apply the two-deficit model to all delusions is overly ambitious. The two-factor cognitive neuropsychological framework is, perhaps, a more realistic model within which to work, although even this model may oversimplify the situation. It remains to be demonstrated whether or not a disturbance in the patient's enormously complex relationship to self, others, or the environment can be adequately explained by the failure of one or two specific bio-mechanical functions. Having said that, I suggest that the two-factor account provides the most explanatorily adequate framework we have to date, and my own account is a two-factor account. Isolating the two factors underlying the formation and maintenance of a monothematic delusion is challenging, but, as will be seen in the next chapter, finding a single factor that accounts for all delusions is even more challenging.

CHAPTER 4: ONE-FACTOR ACCOUNTS

1. Introduction

Some researchers propose that it is only necessary for one factor to be present for a delusion to occur. The most influential one-factor account is that offered by Maher, who considers the presence of an anomalous experience is sufficient to cause a delusion.[16] Although his account is simple and elegant, it is explanatorily inadequate because it fails to explain why patients do not reject a bizarre belief, once formed. The challenge for any single factor account is to explain how a single factor causes both the formation and the maintenance of the delusional belief. Consequently, more recent one-factor accounts are far more complex than Maher's account, as will be seen in the examples that will be discussed in this chapter. Stated simply, for Gerrans the one factor is dopamine (DA) dysregulation. But the disruption it causes impacts on a very complex set of cognitive functions, and the disruption itself varies according to the particular DA process that is disturbed. The one-factor in the Fletcher and colleagues' prediction-error account is aberrant Bayesian inference, but again, the explanation of the connection between this and a delusional belief is quite complex. Because of the complexity of these accounts, the number of factors they evoke is open to interpretation. But notwithstanding the possible disagreement about the number of factors involved, DA dysregulation accounts and prediction error accounts are generally, but not always, regarded as one-factor accounts.

[16] In an early paper on paranoid delusions, and in response to the criticism that it is the readiness of the patient to consider unlikely explanations of her perceptual experience that constitutes the core of the paranoid development, Maher suggests that "as the experiences are likely to be prolonged over considerable periods of time the machinery for reality-testing is itself impaired (Maher: 1974, 108). This implies that a second factor is involved. It appears that he did not pursue this idea as impairment to reality testing is not mentioned in later papers.

2. Phil Gerrans –Dopamine dysregulation

Gerrans offers an account that is more plausible than Maher's, because as well as explaining how delusions are formed, it explains why delusional beliefs are not rejected. His account also differs radically from two-factor accounts in that it does not require a second factor to explain why delusional beliefs are retained. He describes delusion, not as the result of faulty inference, defective hypothesis testing and irrational belief fixation, but as a dopamine (DA) dysregulation that affects the processes involved in constructing an autobiographical narrative response to experience. His account has detail and depth, incorporating neurobiology, phenomenology and evolutionary considerations. According to Gerrans, because the concept of hypersalience described in his account, cuts across the distinction between one and two-factor accounts, there is a sense in which it can be viewed as a one-factor account. However, DA dysregulation impacts on mental time travel (MTT), so to the extent that delusions are a metacognitive phenomenon it may be regarded as a two-factor account.[17] As will be seen in the following discussion, according to Gerrans, both an inappropriate importance given to ordinary objects and events, and impairment to belief evaluation, are caused by the one factor, hypersalience. But because there is a disturbance in both a biological factor (DA dysregulation) and a metacognitive factor (MTT), some people might consider it a two-factor account.

i) Online and offline cognition

Before discussing Gerrans' account of delusions, it is necessary to describe some of the evolutionary, neurobiological and phenomenological considerations that underlie it. Gerrans notes that essentially the mind operates in two modes, online and offline (Gerrans: 2009). Offline processes evolved with the development of the prefrontal cortex (PFC) to "allow us to escape the stimulus-bound present and manipulate representations" (Gerrans: 2007, 35).

[17] Personal communication: Macquarie University 31-07-09.

> The cortical structures which manage the integration and manipulation of representations in offline cognition are layered over phylogenetically ancient systems for the control of behaviour by procedural memory systems. (Gerrans: 2007, 38)

Procedural memory supports habitual behavioural routines such as foraging and navigation. It operates by activating appropriate procedurally regulated memory as the organism encounters various situations. However, when the organism encounters a new situation, it needs to *inhibit* automatic behaviour to allow for the generation of new behaviour, which is then tested and remembered if successful (Gerrans: 2007).

Neural activation in online systems normally only last long enough to control necessary behaviour, but when an unusual situation arises, the activation is sustained long enough to stimulate the provision of additional cognitive resources. The first of these is the focusing of *attention*. Attention and inhibition are correlated because to generate new behaviour attention must be focused and old routines must be inhibited. This is achieved by enhancing attention in the neural assembly activated by the stimuli responsible for the experience and inhibiting attention to competing stimuli (Gerrans: 2007). "A representation that becomes the focus of attention in this way is a *salient experience*" (Gerrans: 2009, 153). According to Gerrans, making representations salient by using mechanisms to selectively inhibit and activate neural circuits is the basic cognitive ability provided by the PFC. "It is a precondition for all the other sophisticated processes in which representations are manipulated" (Gerrans: 2007, 40).

There are three ways offline cognition can be attracted: by a top-down volitional act, by the unfamiliarity of an object or event, or by the stimulation of a posterior representation with a high affect weighting.

> Moment to moment experience is a product of competition between more or less transient coalitions of active neurones, distributed across prefrontal-posterior circuitry, to monopolise offline cognition by attracting prefrontal processing resources. Success in the competition can be produced volitionally, from above, or from below by the novelty or salience of a posterior representation. (Gerrans: 2007, 44)

The connection between the more recently developed PFC and the more primitive posterior areas is bi-directional. While most connections from

posterior networks to the PFC are excitatory, the PFC has extensive inhibitory connection to posterior areas. Thus, the PFC is able to maintain salient representations by enhancing the level of activation in their implementation circuitry and inhibiting activation in competing circuits. The excitatory and inhibitory activities are brought about by the delivery of neurochemicals to neural assemblies. A neurotransmitter in the brainstem binds to receptors at different sites, changing levels of activity in postsynaptic neurons by modulating the response of neurons to other neurotransmitters.[18] Although DA does not act in isolation from other chemical activity, it has been the focus of much research. Gerrans draws on this research, with emphasis on Kapur (2003), to link neurochemical activity to phenomenology. He argues it can be used to explain some characteristics of delusions. Research using neural network models has shown that DA enhances the signal-to-noise ratio (SNR) between communicating neural assemblies. Tonic DA maintains an occurrent activation pattern, thus allowing a representation to be sustained against interference or competition. Phasic DA produces a gating effect, allowing new activation patterns to be formed in the PFC-regulated networks, and the production of salient representations of new stimuli.

> The hypothesis follows, and is confirmed by neural network models, that the balance of tonic and phasic DA is responsible for the role of turnover of representations in the PFC-posterior networks. (Gerrans: 2009, 156)

We can begin to see how this might relate to delusions, as DA dysregulation could disturb the turnover of representations, leading to obsessive thoughts and hypersalience.

ii) Memory and imagination

As Gerrans' account involves autobiographical narrative, it is also necessary to consider the mechanisms responsible for long term memory. Humans are able to encode experiences in retrievable form (episodic memory). Episodic memory is intimately connected with imagination because both can produce perceptual

[18] Neurotransmitters are chemicals released by nerve cells at a synapse for the purpose of relaying information. A synapse is a structure at a point where the plasma membrane of the signal-passing neuron comes into close proximity to the membrane of the target (postsynaptic) neuron.

or sensory imagery in the absence of a stimulus.[19] This enables humans to remember past experiences and imagine possible futures, and to combine different aspects of experience to produce hypothetical scenarios relevant to a decision. The interrelationship between memory and imagination is essential to planning and deliberation. "Offline cognition releases an organism from the stimulus-bound present, allowing us to use information from the recalled past or the imagined future to guide our behaviour" (Gerrans: 2009, 157). The activity of imaginatively projecting oneself into future scenarios and simulating the emotional consequences, which is the direct result of the evolution of the frontal lobes, is called mental time travel (MTT)[20]. The representations manipulated by MTT inherit the affective associations of the online processes. "When you recreate the experience, you must also recreate enough of its online phenomenology to guide deliberation, while inhibiting the rest of its typical online effects such as behaviour" (Gerrans: 2009, 157). As Gerrans points out, "imagining or remembering oneself in a dangerous situation is no guide to action in that situation unless the representation of the experience is characteristically affect-laden" (Gerrans 2007, 41). MTT also involves selective attention to encoding, retrieval, and manipulation of information, in order to provide an autobiographical context for a current experience (Gerrans: 2009).

The representations manipulated in MTT are essentially indexical, being actual or imagined biographical episodes. Therefore, the information they provide is limited. For the purpose of deliberation, we sometimes need non-episodic information from third parties, which they communicate using language. This declarative medium of representation provides access to a far wider range of information than that derived from experience alone. The information, typically facts and propositions, can be stored and recalled as semantic memory. Declarative representations allow us to track our episodic representations, and

[19] For a comprehensive development of this idea see Gerrans and Mulligan *Imagination and default thinking*, 2012

[20] Gerrans stresses that MTT is not the same thing as hypothetical inference, being cognitively prior to it. It is a recreating of autobiographical representations that retain their emotional salience. (Gerrans: 2009).

also to dispense with episodic representations, by allowing linguistic representations to stand in their stead. "The ability to evaluate and act on declarative information *without activating episodic interpretations and their affective associations* is called *decontextualisation* by cognitive psychologists" (Gerrans: 2007, 42). The most abstract form of decontextualisation is the manipulation of uninterpreted symbols**.** The application of inferential rules to uninterpreted symbols in order to solve problem is its most extreme form. Gerrans calls this extreme form 'procedural rationality'. Each level of abstraction away from the original stimulus involves progressively greater inhibitory demands. Going from least to greatest levels of abstraction, these are MTT, declarative cognition and procedural rationality. This brings us to Gerrans' account of delusions.

> Delusions result from the effect on mental time travel of mechanisms which make some representatives hypersalient. That is to say they attract frontal resources, dominating attention, executive function and working memory. Rather than being selectively inhibited, some representations are inappropriately enhanced. Furthermore the very same failures of inhibitory mechanisms which make those representations hypersalient prevent the engagement of more abstract forms of cognition such as decontextualisation or procedural rationality, for these depend on inhibition. (Gerrans: 2007, 43)

iii) DA dysregulation

Gerrans claims that some representations are inappropriately enhanced, and it is this that implicates DA dysregulation. DA alters the explicit (activation-based) and tacit (weight-based[21]) computational properties of the prefrontal-posterior networks by either resetting the weights or by altering the level of activation in a particular neural assembly. Phasic DA, delivered in short bursts, make a target assembly more receptive to increased signals. Tonic DA, which acts longer, makes a target assembly less receptive to signals.

> Offline cognition requires the construction of a transient representation in a coalition of prefrontal-posterior neurones. Whether that representation survives depends on dopamine modulated gating and maintenance effects. Pathological versions of these effects can lead to offline cognition computing over inappropriately maintained representations. Dopamine delivery is thus the mechanism by which representations become hypersalient. (Gerrans: 2007, 50)

[21] "A weight is the propensity, described as a probability, for a unit in a neural network to become active when another to which it is connected becomes active" (Gerrans: 2009, 154).

If the PFC is unable to selectively activate and inhibit representations appropriately, one of two outcomes may arise. Rather than a delusional thought, with the underlying experience that gave rise to it, receding into the cognitive background as it should, more and more cognitive resources are allocated to it until it produces embedded patterns of dysfunctional thoughts and behaviours. Alternatively, a representation may fail to become or maintain appropriate salience, leaving the patient vulnerable to distraction (Gerrans: 2009).

Offline cognition can be described as activation-based processing because it selectively maintains activation in neural assemblies which would otherwise decay.

> Thus, an account of the delusional response to experience is an account of activation-based processing set in train by salient experiences. The difference between delusional and normal subjects lies in the way some experiences (such as the auditory hallucinations of schizophrenia) become salient and how these experiences are dealt with by offline cognition. (Gerrans: 2009, 155)

Normally, when confronted by an experience which captures attention, it is sufficient to consult the episodic database to determine a response. "One remembers or imagines an experience relevant to the current one and assembles a narrative which reconciles them" (Gerrans: 2009, 162). Or alternatively, when presented with theoretical information, most people try to interpret it using their own experience, that is, by consulting episodic memory. The point Gerrans is making is that affect-laden episodic memory is the first step we take in processing that which has gained our attention. Having found a relevant memory, we weave it into our autobiographical narrative in a way that makes sense. If this first step in the process is insufficient, further offline cognitive resources are engaged by inhibiting lower level responses. But if DA dysregulation disrupts the normal gating and maintenance effects that are necessary for the shift to more complex cognition, the experience can acquire undue salience, thus assuming an unwarranted significance. It then begins to dominate offline cognition because the hypersalient experience needs to be incorporated into the autobiographical narrative in a way that makes sense. It is the construction of a narrative around the hypersalient experience that

produces the delusion. This view is supported by the characteristic complex narratives produced by patients with paranoid delusions.

In principle, any thought could become hypersalient as a result of DA dysregulation, but this is not the only way a delusion may come about. Sometimes the delusion arises during the patient's processing of an unusual experience caused by a sensory malfunction in a posterior subsystem, such as an auditory hallucination or a face recognition anomaly in a delusion of misidentification. It is the anomalous experience that commands offline resources because the subject cannot inhibit the problem representation in the normal way (Gerrans: 2007). But again it is the DA dysregulation that is doing all the heavy work because failure of the inhibitory process causes both hypersalience and the failure of more complex offline cognition. Therefore, Gerrans concludes

> DA dysregulation seems more likely to be affecting the processes involved in constructing an autobiographical narrative response to experience than processes involved in a hypothetical explanation of experience. People with delusions are not mad scientists, merely unreliable narrators. (Gerrans: 2009, 165)

iv) Summary and conclusion

According to Gerrans' account, in normal situations neural assemblies are formed which only last long enough to guide action. However, in novel situations the transitory assemblies are sustained long enough to engage offline resources. The first of these is MTT. Past memories are used to interpret the new situation and weave it into autobiographical memory. If this process proves inadequate, further offline resources are engaged. This involves increased distancing from episodic memory (decontextualisation). To enable this process to occur, some neural assemblies need to be sustained and others inhibited. The neural assemblies generated by a novel situation first have to be sustained long enough to engage offline resources, but once offline resources are engaged, the activation of the neural assemblies has to be reduced to allow the higher resources to do their work. The greater the degree of abstraction employed, the greater the inhibitory demands. Research has shown that DA is involved in this

regulatory process of activating and inhibiting neural circuitry. Delusions occur when DA dysregulation causes neural assemblies to be sustained when they should decay or decay when they should be sustained. In the first case, a normal situation becomes hypersalient. The sustained activation engages offline resources, as it should, and using episodic memory, the patient attempts to weave the salient situation into autobiographical narrative in a way that makes sense. When this attempt fails, more abstract offline resources should become involved. However, these resources demand a higher degree of inhibition to enable distancing from episodic memory, and the inappropriate activation of the neural assemblies prevents this from occurring. The patients become trapped at the narrative level, unable to access the higher resources necessary to resolve the anomalous situation, and as it were, to get their story right. In the second case—the premature decaying of neural assemblies—the patients become confused and distracted because of their inability to distinguish between salient and non-salient situations.

Gerrans' account is inspired by Kapur (2003), who proposes that aberrant salience caused by DA dysregulation is implicated in psychosis, and whose theory "provides a heuristic framework for linking the psychological and biological in psychosis" (Kapur: 2003, 13). Gerrans also aims to tell a story that is plausible from an evolutionary, neurobiological and phenomenological perspective. The amount of empirical data he cites, something that is not evident from my brief account, lends credibility to his hypothesis. Unfortunately, the empirical evidence linking DA activity and psychosis, whilst being strongly suggestive, is far from being conclusive. As Kendler and Schnaffner note, the probability of the dopamine hypothesis of schizophrenia (DHS) being true "has declined over the years as non-verifications have substantially outnumbered verifications (Kendler & Schnaffner: 2011, 56). Therefore, the veracity of Gerrans' account remains to be demonstrated. However, support for Gerrans' account may be provided indirectly in a more recent account of schizophrenia, which implicates glutamate dysregulation in the genesis of the condition. As will be seen in the account, which will be

described in the following section, glutamate mediates dopamine activity. As a heuristic account, Gerrans' account gives a plausible explanation of delusions such as paranoia, hallucinations and delusions of reference. However, I am left wondering how his account might be applied to monothematic delusions, such as the Capgras delusion, in which the DA dysregulation would need to be very specifically targeted. Arguably, hypersalience could cause the patient to become trapped at the narrative level, and unable to access the higher resources necessary to correct the false 'impostor' story. What is not clear is why the hypersalience occurs, and only occurs, as a response to the loved-one who is the target of the delusion.

More recently Gerrans (2012) has offered a two-factor account of delusions that focuses on a putative disruption to the inhibition of default thinking necessary to enable transition to the metacognitive level.[22] As for Gerrans default thinking is an imaginative state, the inability to decontextualize leaves the patient trapped in this state. This later offering is far more in keeping with my own two-factor account of the Capgras delusion in which it is hypothesised that monothematic delusions occur when the patient becomes trapped in a fiction.

[22] Gerrans' (2012) two-factor account of delusions is based on his argument that imagination, whilst sharing properties of other mental states, is itself, a distinct mental state, and that our default thinking is imaginative. "Recent work in cognitive neuroscience supports the idea that when human attention is not directed on a stimulus in the environment, cognition reverts to a default mode: basically daydreaming" (Gerrans: 2012, 17). According to Gerrans, the essential properties of imagination are that it is; independent of proximal stimulus, associative, potentially subject to voluntary control, not ultimately responsive to the world and, does not come in degrees. Imagination and imaginative states have these properties essentially. Importantly, imaginative states can play the causal role of beliefs through incorporation. To incorporate is to act on the basis of a mental state without metacognitive evaluation. It is usually unconscious. Gerrans claims most of our cognition is of this kind because metacognitive evaluation is expensive in that it requires inhibition of the default network to enable the decontextualization necessary for evaluating judgments. Delusions occur when the imaginative state is incorporated because of an inability to decontextualize, caused by either hyperactivity or a lesion in the default system. In the Capgras delusion, the first factor is a face processing impairment that triggers the 'impostor' thought. The second factor is the loss of the ability to decontextualize and the subsequent incorporation of the 'impostor' thought.

4. Fletcher and Colleagues' Prediction Error Theory of Delusions

Fletcher and colleagues offer a one-factor hypothesis to explain the symptoms of psychosis. Theirs is a pathological hypothesis, not an aetiological hypothesis, as it deals with the cause of symptoms rather than the cause of the psychosis. It incorporates a prediction error Bayesian model of delusions which subsumes the standard two factors into a single factor, that is, an aberration in Bayesian inference.

> Unlike two-factor theory, our model allows for dysfunctional prediction error to be calculated in PFC and imposed upon the rest of the brain or, alternately for surprise perceptual inputs to arrive at PFC engaging surprise and demanding explanation. Both of these possibilities (bottom-up and top-down) are aberrations of a single factor; Bayesian inference. (Corlett et al.: 2010, 357)

Although they propose that dysfunctional prediction error may result from a top-down or bottom-up dysfunction, most of their work focuses on describing the mechanism involved in the disturbance and transmission of bottom-up prediction error. Their aim is to outline a set of principles that may allow them to "extrapolate from perturbed synaptic function to disordered subjective experience" (Corlett et al.: 2011, 295). The mechanism of top-down dysfunction remains less clearly specified. In their all encompassing model, the challenge is to explain how an aberration at the neuronal level is transmitted up the hierarchy and becomes a delusional belief, and how the predictions calculated in the pre-frontal cortex are transmitted down the hierarchy to the neuronal level. Their explanation involves prediction error learning theory, the concept of the Bayesian brain, and neuropsychopharmacology.

i) Prediction error learning

Prediction error represents the mismatch between what we expect in a given situation and what we actually experience. The mismatch is a signal that our understanding of, or belief about, the world needs to be updated. "By working to reduce this mismatch, we improve our understanding of the causal structure of the world" (Corlett et al.: 2009, 1). If a prediction error cannot be accommodated by the existing world model, new learning and new inferences ensue. Prediction error contributes to learning about rewards and punishments and the predictive validity of information (Murray, Cortlett & Fletcher: 2010).

When a prediction is violated it renders the occurrence worthy of attention and increases its salience, thus converting "the neural representation of an external stimulus from a neutral and cold bit of information into an attractive or aversive entity" (Kapur: 2003, 14). Salience is motivational. It drives thoughts and behaviour because it renders some events more important than others. In tests of associative learning, fMRI results show an absence of the normal neural distinction between important and unimportant events in psychotic patients. This suggests "a role for aberrant incentive salience or dysfunctional prediction error learning in the pathogenesis of psychotic symptoms" (Murray, Corlett & Fletcher: 2010, 466).

The prediction error system is conceived of as a hierarchical system consisting of feedforward and feedback mechanisms. Top-down signalling from neurons in layers higher up the hierarchy confers expectancies on lower levels. Bottom-up inputs to a layer are signalled from the layer below. At any given level, a mismatch between expectancy and input is transmitted up the hierarchy to the level above. The predictions that are conferred on the lower levels are 'best guesses' or 'most probable interpretations' of lower level input, taking into account prior knowledge of the world. The calculated probability being fed down will prove to be accurate, roughly accurate or inaccurate. Any degree of prediction error that occurs is fed forward to higher levels where adjustments can be made to the prediction, and then fed back down the hierarchy.[23] When all available information is incorporated, the prediction and the experience should match. If not, beliefs or understandings need to be updated, and learning occurs. According to Fletcher and Frith (2009), the problem that leads to the positive symptoms of schizophrenia starts with false prediction errors being propagated upwards though the hierarchy. These errors require higher levels of the hierarchy to adjust their models of the world. However, as the errors are false, these adjustments can never fully resolve the problem. As a result, prediction

[23] Some researchers suggest that what is encoded and fed forward is not the raw value but the difference between the raw values and the predicted values. This is obtained by subtracting the predicted values from the actual values. (e.g. Clark: 2012).

errors will be propagated even further up the system to ever-higher levels of abstraction. The severity of the insult may account for how far up the hierarchy a false prediction error will go. Severe insult could create a world in which sensory data are unreliable, making decisions difficult and actions seem fruitless (Fletcher & Frith: 2009).

The concept of prediction error has been formally developed to account for reinforcement and causal learning, the anatomy and physiology of cortical hierarchies and the interaction between basal ganglia learning systems and sensory cortices that mediate perceptual learning (Corlett et al.: 2011). Prediction error is central to formal associative learning theories and it is basic to Fletcher and colleagues' explanation of delusions. A crucial distinction they make is between prediction error per se and the precision or uncertainty about those errors.

> We will develop the argument that delusions (and their neurotransmitter basis) represent a failure to properly encode the precision of prediction and prediction errors: in other words, a failure to optimize uncertainty about sensory information. (Cortlett et al.: 2010, 346-347)

There are two processes in which a lack of precision may prove problematic. One is the lack of precision in the prediction or priors that are transmitted top-down to lower levels. The other is a lack of precision in the encoding of a prediction error transmitted bottom-up to the PFC.

ii) The Bayesian brain

'Bayesian brain' is a term used to refer to the ability of the nervous system to operate in situations of uncertainty in a fashion that is close to the optimal prescribed in Bayesian statistics. [24] Humans operate in a world of sensory uncertainty and our brains must deal with the uncertainty efficiently to generate reliable perceptual interpretations on which to base action. "This leads naturally to the idea that perception is a process of unconscious probabilistic

[24] "According to Bayes rule, the probability that a hypothesis is true given the observed data (the posterior probability) is proportional to the likelihood (the probability of those data given that the hypothesis is true) multiplied by the prior probability (the probability that the hypothesis was correct before the data were seen)" (Cortlett et al.: 2011, 301).

inference" (Knill & Pouget: 2004, 712). The Bayesian approach has proved to be highly successful in designing artificial vision systems and for solving problems associated with machine learning. Such artificial machines have provided a metaphor for neuronal computations (Friston: 2011). The computations that underlie Bayesian inference allow flexibility in a system that needs to deal with sensory uncertainty.

> This allows the system to integrate information efficiently over space and time, to integrate information from different sensory cues and sensory modalities, and to propagate information from one stage of processing to another, without committing too early to particular interpretations. (Knill & Pouget: 2004, 712)

Flexibility is fundamental to the prediction error system hypothesis, according to which a cascade of inferences based on probability occurs when the brain interprets sensory data. But Friston (2011) make the following point.

> It is important to appreciate that the Bayesian brain hypothesis is just a description of optimal behaviour: it does not prescribe how Bayes optimal perception, sensorimotor integration or decision-making under uncertainty emerges. [.....] The Bayesian brain hypothesis just says that perception and decision making are (approximately) Bayes optimal. (Friston: 2011, 3)

Clark (2012) observes that Bayes' optimal response is not a standard that real-world biological systems will typically meet, and it has not been demonstrated that the brain computes probabilities in this way.

> The mere fact that a system's response profile takes a certain shape does not itself demonstrate that the system is computing using a Bayesian apparatus. In a limited domain, a look-up table could (see Maloney and Mamassian (2009), Maloney and Zhang (In Press), yield the same behavioural repertoire as a Bayes' optimal system. (Clark: 2012, 17)

But because so many operations performed by the brain can be described in Bayesian terms, Fletcher and colleagues use it as a unifying standard that serves as a framework on which to base their explanation of the causes of the symptoms of psychosis.

Fletcher and colleagues define beliefs as attitudes we have with regard to propositions about the world. "They are used to predict the future and respond accordingly", thus, "the most rigorous and formal definition of beliefs appeals to probability theory, and in particular Bayesian formulations" (Corlett et al.: 2010, 347).

> In the context of hierarchical Bayesian inference, the posterior belief (having seen the evidence) rests on empirical priors. Empirical prior are prior beliefs that are themselves optimized during hierarchical inference. [...] Assuming that the brain uses hierarchical

> inference to make predictions about the world, most of the beliefs it entertains can be regarded as empirical prior beliefs. (Corlett et al.: 2010, 347)

According to Fletcher and colleagues "beliefs are represented in the brain through the formation and strengthening of synaptic connections between neurons" (Corlett: 2010, 346). The strengthening or extinction of these connections is influenced by the presence or absence of prediction error. When a situation is consistent with a prior belief the neuronal connections strengthen. But if situations repeatedly violate the priors, the synaptic connections weaken to extinction. To explain how an aberration in Bayesian inference can produce a delusional belief, they turn to neuropsychopharmacology.

iii) Neuropsychopharmacology

Fletcher and colleagues suggest that prediction error learning and memory may involve a basic model of brain function called 'predictive coding'.

> Brains, component brain systems and even single neurons minimize uncertainty about incident information (either external or internal) by structurally or functionally embodying a prediction and responding to errors in the accuracy of the prediction. (Corlett et al.: 2010, 348)

According to their model, prediction errors are embodied in neurotransmissions. Bottom-up prediction error signals are transmitted through fast glutamatergic and GABA mechanisms.[25] Slower neuromodulating transmitters, such as dopamine, are engaged in mediating the post prediction error response by encoding the precisions of, or uncertainty associated with a particular prediction error. Dopamine does this by altering the signal-to-noise ratio, thus mediating salience. Uncertainty signals engage subsequent processing, such as enhancing neural maintenance of working memory, and modulating synaptic plasticity down the hierarchy. The sustaining of working memory and the creating of synaptic plasticity enables the modification of neural assemblies. Updating beliefs requires the modification of neural

[25] Glutamate is one of the 20 amino acids, which has the additional function of acting as a neurotransmitter. Glutamate receptors are responsible for the glutamate-mediated postsynaptic excitation of neural cells, and are important for neural communication, memory formation, and learning. Gamma-aminobutyric acid (GABA) acts as an inhibitory transmitter.

assemblies. Therefore synaptic plasticity and the sustaining of working memory are essential to learning. Excessive and inappropriate dopamine signaling is thought to render merely coincidental events highly salient. This may result from a dysfunction in glutamatergic and GABAergic signaling and thence the regulation of dopamine signaling. Inappropriate dopamine signaling stimulates the cascade of events that a true prediction error engages, that is, a search for an explanation that results in new learning.

More specifically, Fletcher and colleagues suggest false prediction error results from a change in signal-to-noise properties of dopamine signaling due to a deficit in glutamatergic regulation of the ventral tegmental area (VTA) dopamine neurons.[26] There are two types of glutamate receptors that are important to their model: the N-Methly-D-aspartic acid (NMDA) and a-amino-3-hydroxy-5-methly-4-isoxazolepropionic acid (AMPA) receptors.[27] Their suggestion that glutamate dysregulation underlies delusion is supported by their research into the effects of ketamine. Ketamine is an NMDA antagonist that binds to the NMDA receptors. It is used as an anaesthetic, and in smaller doses, for recreational purposes. On emergence from anaesthesia, subjects experience psychotomimetic side effects, including illusions, delusions, hallucinations and persecutory thoughts. Thus, ketamine can be used to test cognitive and neural models of how those symptoms arise.

> We can challenge subjects with psychological tasks that engage a number of key cognitive processes; we can capture the neural circuitry engaged by those processes using functional imaging; and, subsequently, we can administer controlled doses of ketamine that induce the psychotomimetic effects of interest. (Corlett et al.: 2011, 296)

They use neuroimaging research to show prediction error signals in the brain can be estimated, even though they do not have any explicit behavioural correlates. Signals have been recorded in the midbrain, striatum and the frontal

[26] The VTA is a group of neurons located close to the floor of the midbrain and is the source of the dopaminergic cell bodies of the mesocorticolimbic dopamine system. It is implicated in the natural reward circuitry of the brain.

[27] NMDA and AMPA are synthetic glutamates that bind to specific glutamate receptors, which are named after them.

cortex[28]. Fletcher and colleagues link the Bayesian brain and neuropharmacology in the following way.

> The Bayesian hierarchy model makes explicit predictions about the roles of different neurotransmitter systems and subsystems in signaling predictions and prediction errors. This division of labor maps on to other conceptions of interneuronal signaling such as the notion of neural 'drivers' and 'modulators'. (Cortlett et al.: 2011, 301)

Priors are set in single neurons by the number and functionality of postsynaptic receptors on each cell, and the tone of slower, neuromodulating inputs. It is not within the scope of this thesis to give a detailed account of the complex interplay of interneurons[29], neurotransmitter systems and subsystems, and other chemical drivers and modulators on which their model is based. The following simplified account is included merely to give the gist of their story.

Changes in neural circuitry result from alteration in the excitatory activity between neurons. Glutamate is the main neurotransmitter in the brain, mediating point to point transmissions across the synaptic cleft in excitatory synapses. The prediction error model is hierarchical, involving feedforward and feedback excitatory activity. This activity is influenced by the activity of interneurons. Whereas principal cells are mostly excitatory, using glutamate as a neurotransmitter, interneurons mostly use GABA to inhibit their target. Interneurons mediate both feedforward and feedback inhibition, but the majority of control involves feedforward inhibition from local interneurons. Striatal GABAergic interneurons control the output of the striatum to the frontal cortex via direct and indirect pathways involving D_1 and D_2 dopamine receptor engagement.[30] There are two subtypes of interneurons that mediate inhibition, fast spiking and low-threshold spiking, reflecting AMPA and NMDA- mediated currents and responses. According to Fletcher and colleagues, this arrangement

[28] The striatum is part of the basal ganglia that receives information from all areas of the cerebral cortex including all forms of sensory information and information regarding the motor system.

[29] Interneurons are neurons that form connections between other neurons.

[30] Dopamine is a neurotransmitter. D_1 and D_2 refer to the type of dopamine that binds at a particular receptor. D_1 receptor signaling and their interaction with NMDA channels facilitate the maintenance of Up states in target neurons and reinforce cell assemblies representing expected salient events. D_2 receptor signaling is associated with instability of prefrontal representations.

would amount to AMPA receptor stimulation driving output, and NMDA receptor signalling inhibiting output, from the striatum to the prefrontal cortex. They make a further speculation.

> We speculate that fast neurotransmitters (GABA and glutamate) may encode the prediction error and slower neuromodulators (eg. Dopamine and acetylchlorine, depending on the task and underlying circuitry) may compute the uncertainty. (Corlett et al.: 2011, 3003)

Using this model, Fletcher and colleagues explain the cause of delusions in the following way.

> In this scheme, the excessive D_2 signaling, impaired D_1 and impoverished NMDA signaling that comprise psychotic states would lead to a poor specification of prior expectancies and fronto-striatal cell assemblies comprised of cells representing merely coincident events and spurious associations. (Cortlett et al.: 2010, 349)

It is by attempting to explain 'merely coincident events' and 'spurious associations' that delusions are formed.

For the above mechanism to work, prior expectancies need to be transmitted top-down to lower levels in the hierarchy. They explain how this might be done using theories of synchrony. One area of this research involves the examination of the role of neural oscillations in inter-area communication. The idea is that high-level cognitive processes cause populations of neurons in different part of the brain to oscillate in the same gamma frequency band. Fletcher and colleagues note that in human learners, gamma oscillation increases during the acquisition of new associations, as does the coherence of oscillations in cortical regions representing the stimuli being associated (Corlett et al.: 2010). They then speculate that the transmission of information between levels in the hierarchy reflect beta-band oscillations.

> While gamma oscillations are commonly enhanced under conditions that involve cognitive control, the top-down specification of priors may reflect in beta-band (15-30 H_Z) oscillations. (Corlett et al.: 2010, 349)

In support of their speculation they note that highly schizotypal subjects have electrocortical responses to sensory stimulation in the gamma and beta frequency range that are slower to habituate following repeated presentation of the stimuli, indicative of maladaptive prior expectancies as well as aberrant error prediction. Further, patients with schizophrenia have reduced long-range phase synchrony in the beta-band during gestalt stimulus perception, which

they suggest, may be indicative of aberrant prediction error (Corlett et al.: 2010).

The above offers an account of the formation of delusional beliefs by suggesting that aberrant prediction error learning leads to false beliefs.

> Essentially we advocate an explanation of delusions as a disruption to the normal Bayesian predictive mechanisms of the brain such that predictable and irrelevant events mismatch with expectancies and their salience demands new learning and explanation; a delusion represents an explanatory mechanism, an attempt to impose order on a disordered perceptual and cognitive world. (Cortlett et al.: 2010, 348)

This leaves the second factor, the incorrigibility of delusional beliefs, to be explained. The Fletcher and colleagues' explanation utilizes theories regarding the strengthening or extinction of neural assemblies in prediction error learning. Research suggests that extinction does not involve unlearning of the original association, but rather, the formation of new associations. "Extinction experiences (the absence of expected reinforcement) invoke an inhibitory learning process which eventually overrides the original cue response in midbrain dopamine neurons" (Corlett et al.: 2009, 3). They note that individuals with psychosis do not learn well from expected events that are absent, nor do they consolidate the learning that does occur. But it remains to be explained why delusions persist in the face of evidence to the contrary. Fletcher and colleagues base their answer on the internal reinforcement hypothesis (IRH). The hypothesis draws on the claim that beliefs are not 'unlearned' but overwritten with new beliefs.

> The IRH contends that extinction and reconsolidation are two parallel processes and that an organism's behaviour is based upon the balance between the strength of their two memory representations following memory retrieval. (Corlett et al.: 2009, 3)

When an expectation is violated, the expected situation is retrieved from memory. So, besides inducing extinction learning, the situation will trigger memories of the reinforced situation. Memories of prior reinforcement can, in themselves, serve as a reinforcer that consolidates the original learning. This can occur, for example, when there is only a small difference between the new situation and the past situations, enabling the violation to be set aside in deference to past learning. Fletcher and colleagues speculate that midbrain dopamine neuron response to conditional stimuli might drive reminder learning

and reconsolidation. As they have conceptualized delusions as a network of associations formed by aberrant prediction error signaling, they speculate further.

> Perhaps we might interpret the persistence of delusions, even in the absence of evidence and the presence of contradictory evidence, by positing inappropriate activity in the midbrain reminder system. That is, aberrant prediction errors might re-evoke the representation of the delusion without definitively disconfirming it. This would drive preferential reconsolidation over and above any new extinction learning. The net effect would be a strengthening rather than a weakening by extinction. (Corlett et al.: 2009, 3)

Consequently, reminding patients of their delusional belief in therapy may increase their conviction. They note there is some evidence that confronting deluded patients with contradictory evidence does indeed strengthen their delusional belief, and that challenging information may be incorporated into the belief.

In summary, Fletcher and colleagues argue that delusion occurs in the context of a noisy nervous system that is attempting to form and maintain a robust set of priors. Excessive noise engenders more cycles of reactivation and subsequent reconsolidation, leading to a bizarre and maladaptive set of expectancies about the world (Corlett et al.: 2009). According to Fletcher and colleagues, beliefs can be viewed as a habitual way of looking at the world, a schema through which future information is interpreted. They suggest that reconsolidation might play a role in developing such a schema. It follows from this, that a delusion might be viewed as a false belief that has become a habit or schema. In support of this model they give such examples as the consolidating of belief in therapy described above, double-bookkeeping, and the persistence of a delusional belief across psychotic episodes. They also note that pharmacological intervention can be substituted for training trials in experimental animals, strengthening the memory trace to the same extent as it is strengthened in actual trials (Corlett et al.: 2009).

iv) Capgras and Fregoli

Fletcher and colleagues begin by noting that the two-factor explanation of the Capgras and Fregoli delusions involves two deficits, one in the perception of affect, the other in belief evaluation.

> Instead, we argue that phenomenology of the percepts are such that bizarre beliefs are inevitable: surprising experiences demand surprising explanations. (Corlett et al.: 2010, 360)

They base their explanation of the Capgras and Fregoli delusions on the speculation that "just like sensory perceptions, emotions are predicted" (Corlett et al.: 2010, 360). It is the prior expectancy that a familiar face will be accompanied by an emotional response that breaks down in the Capgras patient. With Fregoli the problem is misplaced familiarity. They note that right hemisphere damage, commonly the frontal cortex, has been observed in many patients, and that this is in line with their work, which implicates a region of right dorsolateral prefrontal cortex in prediction error signaling. They suggest that this could lead to both an *increase* and *decrease* in perceived familiarity, thus accounting for both Capgras and Fregoli.

Studies of patients with relevant lesions show that they attend to, and learn about, irrelevant stimulus dimensions. So Fletcher and colleagues speculate that damage to the prefrontal cortex might elevate activation in the remaining neurons due to a disturbance in the normal inhibition of subcortical dopamine through glutamatergic and GABAergic mechanisms. The resulting excess in dopamine would create false salience, leading to the Fregoli delusion. The cause of the Capgras delusion is somewhat different.

> Capgras results when patients experience an anomalous lack of affective responding when confronted with the relatives (Ellis and Young: 1990). The delusion constitutes a new prior driven by the experience, a means for explaining it away (Young: 2008). It is possible that the initial affective disturbance results from a failure to guide affect perception by prior experience. (Corlett et al.: 2010, 360)

I find their account of the Capgras delusion unsatisfactory for several reasons. The first is that they do not acknowledge the literature demonstrating that the autonomic response measured in the Capgras delusion, does not register at the conscious level. They base their explanation on Ellis and Young's hypothesis that the delusion is an explanation of the patient's failure to experience the

expected affective response when confronted with their relative. The second is that their account is underspecified. How, or why, the initial affective disturbance results from a failure to guide affect perception by prior experience is not at all clear. More problematic is the number of factors involved in their explanation. They describe two factors: (1) prediction error that causes a false inference, which becomes a belief and (2) disturbance in the midbrain reminder system that prevents the false belief from being over-written. If (2) isn't a second factor, everyone who has the anomalous experience should develop the delusion—and we know that they don't. Arguably, those who have the anomaly but not the delusion are those who able to over-write their false belief generated by the prediction error, and therefore, do not have (2). The Fletcher and colleagues account is either a single factor account that does not explain existing data, or it is a two-factor account.

v) Discussion

The Fletcher and colleagues' model is compelling and impressive in both range and detail. It is also highly speculative. Fletcher and colleagues (2010) acknowledge the speculative nature of their account and stress that the purpose of their model is to provide a framework through which we can build a mechanistic understanding of the puzzling symptoms of psychosis. However, they also claim that their framework offers a single factor account of delusions that subsumes the generally accepted two-factor account. Whilst they may have succeeded in achieving the former, I suggest their argument for a single factor is flawed for the following reasons. Their specific claim is that the two-factor account can be encompassed by a single factor, aberrant Bayesian inference. In doing so they are attempting to subsume two *causal* factors into one *non-causal* factor. To do so is not logically coherent. Further, I question the force of giving a non-causal account of delusions. If we accept that delusions are false beliefs, to say they are aberrant inferences does not advance the debate. It merely re-states the definition of delusions. It does not account for them. Suggesting that the single factor is 'prediction error' does not fare any better. As the non-deluded also make prediction errors, a second factor (such as impairment to a

belief evaluation system) is needed to produce a delusion. Moving down the hierarchy to the level of neuropharmacology does not remedy the problem. Whilst glutamate dysregulation (possibly caused by damage to the prefrontal cortex) might be implicated in the production of both false prediction errors and in the disturbance of the midbrain reminder system, again it cannot be a single factor, because some people have the false prediction but not the delusion. It would need to be explained why a single neurotransmitter dysregulation causes one problem in some people and two problems in others.

A wealth of research remains to be done before a viable causal account of delusions can be produced using the prediction error model. One reason for this is that although the neuropharmacological data on which their model is based is extremely detailed, it is often quite speculative. My brief account does not show the complexity and uncertainty of this data. For example, although they stress the role of dopamine in producing false inference, it is not the only neuromodulating transmitter implicated in the prediction error mechanism. Acetylcholine, serotonin and cannabinoids are also engaged in mediating post prediction error responses (Cortlet et al.: 2010). The glutamate system is also quite complex. For example, it has been observed that dopamine (DA) axons have both synaptic and nonsynaptic terminals. From this observation "the hypothesis has been raised that the synaptic junctions established by DA neurons are the sites at which these neurons release glutamate as a co-transmitter" (Descarries et al.: 2008, 294). The hypothesis is complicated by the fact that other monoamines are also thought to release dopamine as a co-transmitter. More problematic is the fact the purpose of these co-transmissions is a matter of speculation (Descarries et al.: 2008). Further, although 'glutamate spillover' into the extrasynaptic space, and the re-uptake of glutamate by the glial cells[31], is part of the Fletcher and colleague's story, the exact nature of the

[31] Glial cells are non-neuronal cells that supply support and protection for neurons. They modulate neurotransmissions, although the mechanism is not well understood. "The discovery of glial NMDA receptors further indicates the complex nature of intercellular signaling mechanisms in the brain," (Verkhratsky & Kirchhoff: 2007, 28)

extrasynaptic glutamate dynamics in the brain is far from being established. According to Okubo et al. (2010) it has only been *suggested* that glutamate escapes from the synaptic cleft, thus generating extrasynaptic glutamate dynamics. The suggested spillover has been implicated in a variety of important neural and glial functions, of which Okubo et al. list seven.

> Despite the immense potential physiological importance of glutamate spillover, the spatiotemporal dynamics of extrasynaptic glutamate concentration have been only inferred indirectly, and their characteristics remain elusive because of a lack of appropriate technology. (Okubo et al.: 2010, 6536)

In addition, much of this kind of research has been conducted using rats, and it is not known if the data can be applied to the human brain. Nevertheless, Fletcher and colleagues have pieced together a plausible glutamate story using the available information, generally accepted hypotheses and unconfirmed speculations.

Fletcher and colleagues' use of the concept of the Bayesian brain lends coherence to a widely disparate range of theories. It enables them to discuss information regarding the level of single neurons through to higher cognitive processes within the same conceptual framework. But further research is needed here because, as Clarke (2012) notes, to date, no study of prediction error at the single neuronal level has been conducted. Their claim that the two-factor theory can be subsumed into a single factor, an aberration in Bayesian inference, is based on the idea that the same problem that underlies false prediction error (inappropriate dopamine activity in the striatum that results from glutamate dysregulation) also causes the delusional belief to become a habit. Whereas the explanation of the first is detailed and complex, their account of the second amounts to little more than the speculation that midbrain dopamine neuronal activity might drive reminder learning and reconsolidation, and inappropriate activity might lead to a delusional belief moving towards the 'habit' area of the striatum. There is little hard evidence that this is the case.

Fletcher and colleagues cite the following caution offered by Box and Draper (1987). "Remember that all models are wrong, the practical question is how wrong do they have to be to not be useful" (Corlett et al.: 2011, 297). It is

possible that future research may prove wrong many aspects of the Fletcher and colleagues' model, but it is most unlikely that it will be proven so wrong as not to be useful. Their reductionist approach is not in keeping with my holistic approach, and therefore, I do not include their speculations in my account. Having said that, both my delayed-processing account of the first factor and my coherence-monitoring account of the second factor, could be construed in terms of prediction error.

CHAPTER 5: TWO CONTENTIOUS ISSUES

1. Introduction

There are two contentious issues relating to monothematic delusions that have yet to be resolved. The first is their doxastic status. The DSM defines delusions as beliefs, but some authors, such as Berrios (1991), Currie (2000) and Sass (2001), challenge this definition. They argue that delusions are not beliefs. The second issue concerns their perceptual content.

> Some accounts of phenomenal consciousness entail that high-level content is phenomenally inadmissible, other accounts leave the possibility of high-level phenomenal content open, and still other accounts require that high-level representations are phenomenally admissible. (Bayne: 2009, 386)

The poverty or richness of the perceptual content is important to accounts of monothematic delusions because any gap between the perceptual content and the belief needs to be adequately explained. Further, the size of the gap determines the ease with which the incorrigibility and subjective conviction of delusions can be explained. In the first section I will discuss the doxastic status of delusions and the strengths and weaknesses of the different views. In the second section I will discuss various arguments for thin and thick perceptual content, concluding with an explanation of my own position.

Section One: EXPERIENCE AND BELIEFS

1. Introduction

Following Maher, most accounts of delusions propose that delusions are beliefs formed to explain an unusual experience. In keeping with Jaspers, others hold that the unusual experience *is* the delusion, which may or may not give rise to a belief, whilst still others hold either delusions are never beliefs or that they are not based on experience. In keeping with the DSM definition, the doxastic status of delusions is widely accepted, but the role experience plays in their formation is still a matter of debate.

> The precise role of the experiential component in the pathogenesis of monothematic delusions, such as delusions of alien control, is not yet understood. Not only is it extremely difficult to test various hypothesis in experimental conditions, most hypotheses in the area also invoke deep and poorly understood conceptual issues concerning rationality, experience and belief-formation in normal subjects. (Hohwy & Rosenberg: 2005; 142)

The doxastic accounts can be broadly divided into two approaches, the endorsement and the explanationist, although a recent account places the two approaches at the extreme ends of a continuum (Langdon & Bayne: 2010) and a second blurs the distinction (Coltheart, Menzies & Sutton: 2010). This chapter will begin by briefly acknowledging three influential non-doxastic accounts: those of Currie, Berrios, and Sass. The difference between two doxastic accounts, the explanationist and the endorsement accounts, then will be given, followed by a brief description of the new accounts offered by Langdon & Bayne and by Coltheart, and a discussion of their attempts to resolve the differences between the endorsement and explanationist accounts.

2. Non-doxastic accounts

There are two overriding difficulties facing those wishing to debate the doxastic/non-doxastic status of delusions. There is a lack of consensus on the criteria a belief has to meet to qualify as a genuine belief, and there is a lack of agreement regarding the processes involved in the formation of normal beliefs. Although this clouding of the issue is true of all accounts of delusions, it is particularly relevant to the doxastic/non-doxastic debate, much of which is fundamentally a disagreement over one or both of the above issues.[32] Having said that, there are four characteristics generally held to be necessary for a belief; a) beliefs are typically manifest in behaviour b) beliefs can be self ascribed and defended with reasons c) beliefs are sensitive to evidence and argument d) beliefs are integrated in a web of other beliefs and the subject's self narrative. Failure to comply with one or more of these characteristics results in the mental state being disqualified as a belief. However, Bortolotti suggests that the main problem with objections to the doxastic status of beliefs is that they "assume an idealised conception of beliefs, and blind us to the possibility that

[32] For a full account of this debate see *Delusions and other irrational beliefs* (Bortolotti: 2010)

we are all procedurally, epistemically and agentially irrational in systematic and (sometimes) disturbing ways" (Bortolloti: 2010, 59).

i) Berrios

Berrios (1991) questions whether delusions have any genuine content and suggests they are empty speech-acts that disguise themselves as beliefs. As Davies and Coltheart (2000) describe this view, the utterances should be regarded as expressions of the subject's thought but, from a psychological view, as nothing more than 'mere noise'. Berrios holds that delusions are not beliefs because they are not contentful states. What appears to be content refers neither to the world nor the self. Patients may be able to emit a verbal formula but are unable to say what the words mean or to discuss their implications. Further, propositions so clearly falsified by other facts available to the patient cannot be sincerely believed. His argument for delusions being non-doxastic is based on their apparent failure to comply with b) and c), that is, they cannot be ascribed to the self nor defended with reasons, and they are not sensitive to evidence or argument. Young (1999) argues against this view by placing more emphasis on a) and arguing that sometimes patients do act on their delusions, which shows they are not just empty speech-acts.

ii) Currie

Currie argues that delusions are not beliefs but imaginings that are mistaken for beliefs (Currie: 2000; Currie & Jureidini: 2001). According to his theory, the schizophrenic patient has a problem distinguishing between merely imagining some proposition and really believing it. He suggests that if we viewed the thoughts of schizophrenic patients as a flow of imaginings we would not consider them strange or bizarre. Currie says we hold that representational states are beliefs in virtue of their occupancy of certain functional roles, which includes roles in theoretical inference, practical reasoning and generation of action. He argues that although imaginings are significantly similar to beliefs, they are not the same. For example, despite exerting a powerful psychological

force, imagination fails to engage action in the way a genuine belief would. Patients often fail to act on their delusion and this indicates they are more likely to be imaginings than delusions. To account for those who do, he says because the patient believes she believes that P, and there is a general assumption that one's beliefs are true, the patient can come to believe that P and act on it. Davies and Coltheart (2000) note that this means Currie's theory still allows that some schizophrenic patients have delusional beliefs. According to Currie, the cause of the confusion between imagining and beliefs is a general loss of a sense of agency, thus failing to comply with b). He bases this suggestion on Frith's theory of efferent copying.[33] According to Currie (2000), if efferent copying fails, thus causing a loss of the sense of agency, people will formulate a thought but not recognise it as a thought of their own. Currie's attempt to apply the efferent copying theory of motor action to thought is a weakness in his account, because the command to have a thought, on which the prediction is based is, itself, a thought and it leads to an infinite regression. A further weakness is that it is not clear how a general failure in agency, or a general confusion between imaginings and beliefs, could explain monothematic delusions.

iii) Sass

Sass views delusions as an alteration in the way the patient experiences her world. Therefore, he claims the beliefs resulting from these states are not normal beliefs. He suggests the inadequacy of traditional views lies in certain concepts, such as 'poor reality-testing'.

> The official definitions of these concepts all imply that, while the *content* of such experiences is incorrect, their *form* or ontological status, the way in which the patient experiences what he experiences, is identical to that of actual perception and accurate beliefs. (Sass: 1992, 113)

This makes delusions beliefs in the normal sense, but ones that are false. Sass argues that several features of schizophrenic behaviour suggest delusions are

[33] Frith's theory of efferent copying proposes that prior to a motor action the brain generates an efferent copy of the actual motor command and uses it to make predictions about the effects of the action. The predicted sensory effects of the efferent copy are then compared with the actual sensory effects of the motor action. If there is a mismatch, the sensation is labelled as being externally generated (e.g. Frith & Done: 1988; Frith, Blakemore & Wolpert: 2000; Frith: 2004).

not normal beliefs. For example, the degree of indifference to argument and counter-argument seem to distinguish delusions from other beliefs, since they seem in principle to be divorced from any possible need for corroboration. Further, the failure of the patient to act on the basis of these experiences, thus failing to comply with a), suggests that "this is a domain where reality-testing simply plays no role, a domain to which neither belief nor disbelief of the usual kind is of any relevance" (Sass: 1992, 115). Sass contends that the content of the delusion is better understood as an expression of a disturbance in the patient's general mode of experiencing her world than as the kind of mistake that occurs within the common-sense world. He admits to some kind of quasi-belief which is based on the patient's experience, but he does not accept them as true beliefs. For Sass the delusions is not the belief but the altered experience. My account is in keeping with Sass in that it hypothesises that delusion results when the meaning component of the immediate perceptual experience is drawn from an alternative reality, rather than merely being due to a mistake made in the experience of everyday reality. There is an alteration in the patient's usual mode of experiencing her world.

3. Doxastic accounts: explanationist v/s endorsement

In keeping with the DSM, it is generally accepted that delusions are beliefs. However, there are two different ways of describing the way in which delusional beliefs are formed, namely, by *endorsement* or by *explanation.* According to explanationist accounts, delusions are formed by attempting to explain an unusual experience, whereas according to the endorsement account, delusions are formed by endorsing an unusual experience, that is, by accepting the experience as veridical. A significant difference between the two accounts relates to the role experience plays in the formation of the delusional belief. Experience plays a more direct role in the fixing of the delusional content in endorsement accounts than it does in explanationist accounts. Because the experience is taken as veridical in the endorsement account, the content of the perceptual experience has to provide the content of the belief. Therefore, the perceptual content needs to be rich. For example, in explanationist accounts of

the Capgras delusion, the content of the experience generally is held to be non-specific, something like 'there is something odd about my wife', and the content of the belief is acquired by way of explanation. In contrast, the endorsement account attributes a richer content in which the patient experiences his wife as a stranger or as an impostor. Briefly, the difference between the two accounts is that one is an *explanation* of an abnormal perceptual experience with thin content, whilst the other is an *endorsement* of an experience with rich content.

i) The explanationist account

The strength of the explanationist account is that the content of the experience is viewed as relatively thin and non-specific.[34] This is a strength because it is much easier to account for thin content, such as 'there is something odd about my wife' or 'that woman does not feel like my wife', rather than for the thick content, 'that woman is an impostor'. "It is relatively unproblematic to think that perceptual experience might be able to convey the vague content 'something is wrong'" (Langdon & Bayne: 2010, 333). But the thin content has a disadvantage. More work is needed to explain how the thin experiential content gives rise to the content of the delusional belief. "The raison d'être of explanationist accounts is to fill the gap between the contents of the experience and the content of the delusional belief" (Pacherie: 2009; 119). Filling the gap is by no means an easy task. Fine, Cragie and Gold (2005a; 2005b) describe the following challenges explanationists face when attempting to do so. The first is to show that the patient's explanation of the anomalous experience is a good one. Using the Capgras delusion as an example, they suggest there are better explanations for the anomalous thought 'there is something odd about my wife' than that the wife is an impostor. Addressing the proposition that the feeling of oddness is caused by a lack of affect, Fine *et al.* (2005a) note that the impostor hypothesis does not explain why there *is* something odd about her. If the wife was a

[34] The following comments on the explanationist account do not apply to Coltheart, Menzies & Sutton's (2010) account, which blurs the difference between the explanationist and endorsement accounts by placing the production of the rich 'stranger' content wholly at the subpersonal level. Coltheart's account will be discussed later in this chapter.

convincing impostor there would be nothing odd about her and she should elicit the normal emotional response. She would neither *seem* odd, nor *feel* odd.

> If someone were perceptually indistinguishable from a loved one, it is no explanation at all of the absence of emotional responsiveness to say that they are a duplicate. (Fine et al.: 2005b, 147)

Thus, the impostor hypothesis does not explain the experience of reduced affect that is often thought to underlie the Capgras delusion. In fact, "The explanation seems to miss the mark entirely" (Fine *et al.* 2005a; 160).

They note a second challenge for explanationists is to determine the precise content of the experience that needs explaining. Again, the Capgras delusion provides an example of such challenge. Fine, Cragie and Gold (2005b) claim there are no empirical grounds for assuming that it is a change in the feeling of *familiarity* of the person being regarded that the patient is trying to explain. The fact that prosopagnosia patients record heightened autonomic responses to previously familiar faces they no longer overtly recognise nor rate as familiar, indicates that such responses do not register at the conscious level. This suggests that there is not the association between autonomic responsiveness to faces and the subjective feeling of familiarity the explanationist account seems to require. Hohwy (2004) also notes that the skin conductance response does not explain the character of the subjective experience that is hypothesised to give rise to the delusional belief, because it is not part of the experience itself. "Something more is required to give a full account of the type of experience that gives rise to the delusion" (Hohwy: 2004, 66). The fact that patients are typically at a loss to explain what makes them think their loved one is an impostor supports this argument. When pushed, patients might describe small differences, not all of which refer to their appearance (Pacherie: 2009), and these are likely to be confabulations.[35] If the anomalous experience is 'the person looks familiar but doesn't feel familiar', (Stone & Young: 1997), it could be expected that this would be reflected in the explanations patients offer. That this is not the case supports the hypothesis that the discrepancy in the

[35] Examples Pacherie gives are changes in behaviour (the way the impostor ties his shoe laces) and changes in personality (Pacherie: 2009: 117, 114).

autonomic response registers at the subconscious level but not at the conscious level. But then as Hohwy (2004) says, what is it in the experience the patient is trying to explain if it is not the loss of the feeling of familiarity?

A third challenge for explanationists is justifying the conception of the content of the delusional belief as an 'explanation'. Fine *et al.* (2005a) suggest there are two senses in which the explanations patients offer fail to be 'explanations'. The first is that they are not obviously empirically adequate. The second is that they are not just incorrect but so bizarre that they strain the very notion of explanation. They are nothing like explanations as we understand them. This is true, not only of the 'impostor' explanation of Capgras patients, but also of explanations of anomalous experiences such as the belief one is dead (Cotard delusion), that strangers are really friends in disguise (Fregoli delusion) or that one's arm or leg belongs to someone else (somatoparaphrenia). Explanationists attempt to defend the 'explanation' status of delusional beliefs by drawing a logical connection between a putative underlying impairment, the thin experiential content and the content of the belief. It is not clear that they have succeeded in their endeavour, because for many monothematic delusions, such as Cotard and somatoparaphrenia, a logical connection is difficult to draw. For example, the logical connection between the lack of sensation in a paralysed limb and the belief that the limb belongs to another person has yet to be explained. Other delusions, such as delusions of alien control, appear to fare better because the putative link between the neuropsychological impairment and the content of the delusional belief seems more direct. But according to Spence (2001), the phenomena of alien control, by their very nature, implicate 'others'—those allegedly exerting control. To generate the explanation that provides the content of the delusional belief, the experiential content would need to be something like 'my thoughts/actions/feelings *are not my own'*, not merely a vague feeling something is wrong. Frith's theory of efferent copying appears to address this problem with regard to alien control of motor action in that a mismatch between predicted sensory effects and actual effects causes an action to be labelled as externally generated. But alien control of thoughts and of

feelings remain to be explained. “We still do not understand the origin of the most mysterious symptom of all: thought insertion” (Frith: 2012)

A fifth challenge for explanationists is to explain why all patients form and accept the same hypothesis for any particular anomalous experience when the thin content allows for a variety of explanations. Coltheart’s more recent account attempts to meet this challenge by placing the anomaly at the subpersonal level, and demonstrating, using Bayesian principles, how the same explanatory hypothesis would always be selected. But as will be seen later in this section, Coltheart’s account blurs the distinction between the explanationist and endorsement accounts.

Incorrigibility is a further challenge for the explanationists because one would think that an explanation, in the sense that ‘explanation’ is normally understood, would be responsive to reasoned argument and counter evidence. Although some progress can be made with Capgras patients using Cognitive Behavioural Therapy (CBT[36]), the gains tend to dissipate when CBT is discontinued. The subjective conviction that underlies the incorrigibility is also a problem for the explanationists account because it is better explained by the endorsement account. Experience provides more compelling reason for belief than does an hypothesis, and the beliefs experience generates are more resistant to correction.

Despite the difficulties the explanationist account faces, it is not without merit. Directly linking the content of a particular delusion to a specific neurological impairment has explanatory power. The problem may lie, not in the account itself, but in our present inability to determine the precise nature of the impairment and content of each specific delusion. Whilst the precise locus and nature of the impairment, and the richness or poverty of the perceptual experience are still matters of debate, any explanationist account of a particular

[36] CBT employs evidence to challenge dysfunctional beliefs. When used to treat hallucinations and delusions it employs persistent gentle discussions about the evidence for the belief.

delusion remains largely hypothetical. Future research may resolve these difficulties and give further support for explanationist accounts.

ii) The endorsement account

The endorsement account is strong precisely where the explanationist account is weak. By viewing the contents of the delusional belief as an expression of the anomalous experience, it is much easier to give a good account of the subjective conviction and incorrigibility of the delusional belief. As Langdon and Bayne (2010) note, perceived content has a special status. When the content is automatically uploaded as a belief, it renders the belief resistant to revision. Believing what we see without reflecting upon it is our default position and most of our perceptual experiences are of this kind. As we interact with our world, we merely take our experiences as veridical unless we have good reason to do otherwise. We believe the sky is blue with an incorrigible, subjective conviction because we perceive it as blue, not because we have a convincing argument for it being blue. If indeed delusions are expressions of anomalous experiences which are taken as veridical, it gives a good account of their characteristic incorrigibility and lack of reason.

The natural tendency to upload an experience to belief is an 'hypothesised competence' known as the pre-potent doxastic response (Hohwy & Rosenberg: 2005). The response takes experiences and turns them into beliefs. However, if the experience does not fit well with the person's existing web of beliefs, or when some other internal discrepancy is detected, it is hypothesised that the pre-potent doxastic response is inhibited to enable the experience to be evaluated before it is uploaded as a belief. It is further hypothesised that delusion occurs when there is a failure to inhibit the pre-potent doxastic response when there is good reason to do so. Consequently, the anomalous experience is taken as veridical, uploaded as a belief, and then held with the subjective conviction and resistance to revision characteristic of beliefs that are formed in this way. If delusions are caused by a failure to inhibit the pre-potent doxastic response, it needs to be explained why patients only develop delusions

in specific situations and not every time their senses trick them. In answer, Hohwy and Rosenberg (2005) hypothesise that the lack of inhibition of the pre-potent doxastic response is a localised reality testing performance failure that occurs in specific circumstances, rather than being a general competence failure. They focus on the reality testing that follows any observed mismatch between experience and background beliefs. The intra- and inter-modal reality testing, such as listening to a sound more closely, or looking for something we thought we heard drop, can inhibit the pre-potent doxastic response. However, "When such reality testing procedures are exhausted, nothing else will, on its own, inhibit the pre-potent response" (Hohwy & Rosenberg: 2005, 149). They point out that not all types of experiences are equally accessible to the normal procedures of reality testing. For example, emotions and affect cannot be tested using a different modality. We cannot test how we feel by looking more closely. This is important because they claim that in some cases there is no other avenue for reality testing other than intra-or inter-modal testing. They suggest purely theoretical belief is non-experiential and, therefore, not a relevant source of reality testing, that another person's testament is only plausible if she is in a better position to evaluate the experience, and background beliefs may not be strong enough to outweigh a compelling experience. They conclude

> Unusual beliefs arise when unusual experiences are taken as veridical because they occur in sensory modalities or at processing stages where application of the available reality testing procedures keeps giving the same result and where further intra-or inter-modal reality testing cannot be performed. These performance failures happen even though the patient does not have a deficit of reality testing competence, that is, even though the patient is able to inhibit the pre-potent doxastic response for other experiential content. (Hohwy & Rosenberg: 2005, 153)

The hypothesised failure of the autonomic response in Capgras patients is an example of an anomaly at a processing stage that is inaccessible to further reality testing. Without any method for further testing, "it may simply not be subjectively possible to inhibit the pre-potent response so as to dismiss the belief or suspend judgment" (Hohwy & Rosenberg: 2005, 155).

The incorrigibility of delusional beliefs is further explained by the normal tendency people have to favour first-hand over second-hand or third-hand sources. If the Capgras patient, for example, has *experienced* the loved-one as an

impostor, this will carry far more weight than the fact others claim that they do not have the same experience. A further observation made by Langdon and Bayne (2010) is that the patient who accepts an experience as veridical will have no sense of themselves as adopting the delusional belief for explanatory reasons, whereas the patient who arrives at the belief by reflecting upon a thin experience might have some awareness of the belief as an explanation. Any awareness of the content of the belief as an explanation provides an opening for rational argument based on counterevidence, whereas a complete lack of awareness leaves no such opening.[37] Therefore, the incorrigibility of delusions and their resistance to counterevidence and rational argument is more in keeping with the endorsement account.

The endorsement account's thick experiential content places restraints on the content of the belief, thus making it easier to account for the common themes that manifest in the content of delusional beliefs. If for example, the Capgras patient experiences his wife as an impostor, there is very little room for him to form a belief other than that she is an impostor. However, not all delusions fare well in an endorsement account. Persecutory delusions are particularly problematic. The problem can be seen by taking Maher's (1974) example of the elderly who are unaware they are suffering hearing loss. Phenomenologically an elderly person may experience a gradual diminution in loudness with which others are speaking. This is easily interpreted as whispering. The sufferer then needs to explain why other people are whispering and why they deny it when questioned on the matter. From this it is easy to conclude that an attempt is being made to conceal something from them, "thus a delusion of conspiracy is not only possible but eminently reasonable" (Maher: 1974, 106). It is difficult to see how this type of delusion could be formed by accepting the experience, in this case of people whispering, as veridical. Some level of explanation seems to

[37] Again, this argument does not apply to Coltheart, Menzies & Sutton's (2010) account in which the content of the delusion is described as being produced entirely at the subpersonal level and passed to consciousness fully formed.

be required to fill the gap between the content of the experience and the 'conspiracy' content of the delusion.

If the raison d'être of the explanationist accounts is to fill the gap between the content of the experience and the contents of the delusional belief then the raison d'être of endorsement supporters is to demonstrate that there is no gap, thus making the explanation account superfluous (Pacherie: 2009). The main weakness of the endorsement account becomes evident at this point. The thick experiential content is at the same time both strength and weakness because it is not at all clear how the experience can acquire the thick content required to close the gap. The failure of the endorsement supporters to give a persuasive account of the thick content of the experience is the primary reason explanationists give for dismissing the endorsement account as inadequate. This is a weakness I will attempt to address in the second section of this thesis, in which I describe how thick experiential content might be acquired.

4. Two Ways of Resolving the Differences

Two recent accounts have attempted to harness the strengths of both the explanationist and endorsement accounts whilst avoiding their weaknesses. The first is Coltheart's explanationist account described in Chapter 3. The second is Langdon and Bayne's (2010) continuum theory.

i) Max Coltheart

Coltheart, Menzies & Sutton's (2010) new account radically departs from Maher's original explanationist account in that it no longer describes the belief as being an explanation of an anomalous *experience*. Having observed that in many monothematic delusions, such as the Capgras delusion, something abnormal occurs of which the patient is not conscious, he places the entire process at the unconscious level. Thus, for Coltheart, the belief becomes an explanation of anomalous *data*. According to Coltheart, subpersonal systems generate suitable hypothesis, which, if true, would explain the actual state of

affairs caused by the anomalous data, then based on probability, selects the hypothesis that best suits the observed data. Using the Capgras delusion as his example, he argues that by a process of abductive inference, of the two generated hypotheses 'that woman is a stranger' and 'that woman is my wife', the 'stranger' hypothesis would be selected in preference to the 'wife' hypothesis. Coltheart argues the point in the following way.

> An explanatorily adequate abductive inference ..... is that the reason this person's face no longer evokes a strong autonomic response is that the person is a stranger, no matter how much this person looks like one's spouse. This is explanatorily adequate because, if it were the case that the person was a stranger, the sight of that person would indeed not generate a strong autonomic response in the patient. (Coltheart: 2007, 1048)

The selected hypothesis would then be accepted as a belief and conveyed to consciousness fully formed. Experience plays no part in forming the content of the delusional belief. "What's conscious is only the *outcome* that this chain of processes generates: the belief 'this isn't my wife'" (Coltheart, Menzies & Sutton: 2010, 264). Coltheart's account increases the thickness of the perceptual content from 'there is something odd about my wife' to 'that woman is a stranger'. By describing the process of belief formation as entirely unconscious, his account gains the strength of the endorsement account because it closes the explanatory gap between conceptual content and the belief. Also, any beliefs formed at the subconscious level acquire the characteristics of subjective certainty and incorrigibility, because processes at that level are not accessible to reason. At the same time it remains an explanationist account, because the belief results from the forming of hypotheses designed to explain the anomalous data. Extrapolating from the last point, all endorsement accounts involve elements of explanation in that the meaning attached to the immediate perceptual experience is the result of 'inference to the best explanation', so there is a sense in which all endorsement accounts could be regarded as explanationist accounts. Alternately, we might consider Coltheart's account to be an endorsement account, because the meaning of the perceptual data is processed at the subpersonal level, taken as veridical and uploaded as a belief. Thus, Coltheart blurs the distinction between the two accounts.

ii) Langdon and Bayne

Taking an inclusive approach, Langdon and Bayne (2010) describe a spectrum of delusions along a dimension they label 'received-reflective'. They propose that all delusions lie somewhere on this spectrum with the content of each delusion formed by a variable weighting of endorsement and explanation processes.

> We propose that both accounts apply to the explanation of delusions, and that all delusions lie somewhere on a received-reflective spectrum. At one pole of this spectrum are "received" delusions formed solely via "endorsement processes" (i.e. those processes of delusion formation according to an endorsement account); at the other pole are "reflective" delusions, formed solely via "explanationist processes". (Langdon & Bayne: 2010, 320)

To accommodate the endorsement component of the spectrum they adopt a reasonably liberal conception of the admissible contents of experience, according to which "a fairly broad array of high-level properties can be experientially encoded" (Langdon & Bayne: 2010, 333).

The similarity they see between the endorsement and explanationist routes to belief formation is that they both involve inferential transitions, with the difference being in the placement of the transition in relation to experience.

> The explanationist account of the route from experience to a reflective delusion locates most of the inferential processing downstream of experience, in the process of hypothesis generation and evaluation, whereas the endorsement account locates it upstream of experience, before the received delusion is endorsed. (Langdon & Bayne: 2010, 331)

They flesh this out by giving two accounts of the formation of the Capgras delusion. In the first (explanationist), the experience has thin content, such as 'there is something wrong about that person', and the impostor hypothesis is acquired by reflecting on the experience. In the second (endorsement), the content of the experience is 'that person is a stranger'.

> This experience of seeing a stranger is then accepted as veridical and the patient believes, on seeing the stranger, that the person in front of him or her, *is* a stranger. The Capgras content is then but a simple – and perhaps spontaneous and unconscious – inference away. (Langdon & Bayne: 2010, 328)

If the experience is 'stranger', the inference that produces the Capgras content (impostor), whether made consciously or not, is downstream, not upstream of the experience. Therefore their endorsement account allows for some post-experiential inference.

Langdon and Bayne propose that most delusions are hybrid in that they arise from a combination of endorsement and explanationist processes. Different delusions are formed by different weighting of received and reflective components, with some delusions being better explained by a weighting more towards the received end of the spectrum and others more towards the reflective end. Persecutory delusions are viewed as lying closer to the reflective end of the spectrum because it is assumed that it takes considerable reflection over time for a suspicion to evolve into a full blown delusional belief. On the other hand, mirrored-self misidentification lies closer to the received end of the spectrum. If the patient suffers from mirror agnosia and sees the mirror as a window, the content of the experience (that person is a stranger) requires little reflection. The weighting of other delusions is less clear and they offer two interpretations of the Capgras delusion, one weighted towards the received end of the spectrum and the other towards the reflective end of the spectrum. In doing so, they note they have no principled objection to the proposal that a particular delusion in one patient might be relatively more received whereas the very same delusion in another patient might be reflective.

Langdon and Bayne favour a two-factor account of delusions and propose that the second factor development will be constrained by one's conception of the role of experience in relation to the first factor which explains the delusional theme. If for example, the Capgras delusion is viewed as having thick content and therefore lying closer to the received end of the spectrum, the second factor is more likely to be a failure to inhibit the pre-potent doxastic response. Alternatively, if the experience is viewed as lying closer to the reflective end of the spectrum, biases such as motivational bias or external attributional style will feature in the second factor. They acknowledge that in the case of a delusion being a hybrid, two-factor theorists will need to consider both the loss of inhibitory control and the possibility of a disruption to post-reflective inferential processes.

The Langdon and Bayne spectrum account has initial plausibility because it appears to harness the strengths of both the explanationist and endorsement accounts. But their account is not without problems. As well as gaining the strengths of both accounts it also inherits their weaknesses. Explanationists will not accept the thick experiential content necessary for the received end of the spectrum without a convincing argument, and no such argument is given. Arguing the case is outside the scope of their paper, but it leaves the main criticism of the endorsement account unchallenged. At the other end of the spectrum, reflective delusions, such as persecutory delusions, are held with a subjective conviction and an incorrigibility that has not been explained. The hybrid delusions that lie somewhere between the extremes fare better because they do not require such thick experiential content as do the purely received delusions, thus avoiding the thick content problem, whilst at the same time acquiring some of the subjective conviction and incorrigibility of experiences that are taken as veridical. Perhaps the greatest problem for their account is the question of empiricism. Langdon and Bayne say "where a particular delusion might fall on the received-reflective spectrum is an empirical matter that must be decided on a case-by-case basis" (Langdon & Bayne: 2010, 331). It is not clear how empiricism can play a role here when the thickness or thinness of the experiential content remains a matter of philosophical debate, and where the position a delusion occupies on the spectrum depends on one's particular conception of experiential content. A further complication is that according to Langdon and Bayne, the very same delusion could result from different received/reflective weightings in different patients. That we can tell two different stories for the same delusion raises two issues. The first is whether or not two delusions formed in different ways can be regarded as the same delusion. The second is the question of how it is possible to distinguish between the two empirically. Although both stories are theoretically possible, fine-grained neuropsychological data are needed to prove the case, and to date such data are far from being available. The Langdon and Bayne account is thought provoking but unconvincing.

5. Conclusion

From the above discussion it can be seen that both conceiving of delusions as beliefs formed by explaining experiences and conceiving of them as beliefs formed by endorsing experiences, have advantages and disadvantages. Resolving the contentious issues involved is seriously hampered by our lack of knowledge regarding the nature and formation of normal beliefs. One of the specific issues involved in the debate concerns the poverty or richness of the content of the perceptual experience. As discussed above, if the content is thin, endorsing the experience will not provide the full content of the delusional belief. Some post-experiential inference would need to be made to bridge the gap between the content of the experience and the content of the delusional belief. On the other hand, if the perceptual content is extremely rich, a disturbed perceptual experience may be all that is necessary to produce a delusion. The views of three important contributors to the debate regarding perceptual content will be discussed in the next section.

Section two: THE CONTENT OF PERCEPTION

1. Introduction

The most common criticism of the endorsement account of delusions is that it cannot adequately explain how the richness of perceptual content, which the account requires, is acquired. According to the endorsement account, a delusional belief is acquired by taking an anomalous experience as veridical. Therefore, the content of the perceptual experience has to be rich enough to contain all, or nearly all of the content of the belief. But critics of the account claim that it is not possible for the perceptual experience to contain rich content such as 'that woman is a stranger pretending to be my wife', or 'someone is putting thoughts into my mind'. In this chapter I will argue that this criticism only makes sense if 'perception' is restricted to 'sensory perception' isolated from context (e.g. visual face processing in the Capgras delusion). If perception

is taken in a broader sense, in which objects are perceived in context, the criticism does not hold.

2. Defining Perception and its Contents

How one defines perception is largely dependent on the theoretical position one holds, and there is a wide discrepancy in theoretical positions. Historically, a division has been made between *sensation* on the one hand and *judgment* on the other. Jaspers reflects this view when he says that purely sensory features are, "so to speak, brought to life by" awareness of meaning or 'thought' (Jaspers: 1963, 60), and that "we have to distinguish between the original *experience* and the *judgment* based on it (Jaspers: 1963, 96). Another way of describing the divide is to say that there are perceptual mechanisms that process raw sensory data on the one hand, and on the other, cognition which interprets the output of the processing in a way that adds meaning. However, the issue is not clear cut because some theorists blur the boundaries. For example, when Fodor differentiates between the encapsulated, periphery modules that process sense data and the central processing system that performs higher cognitive functions, he includes basic categorizations in the former. According to the traditional view, forming a judgment that a perceived shape is in fact a dog, for example, would fall on the cognitive side of the divide. Churchland blurs the distinction further by suggesting there is no theory-neutral, original experience on which we pass judgment because all levels of perception are open to top-down influence. In contrast, Raftopoulos restates the traditional distinction between processing sense data and making judgments when he draws the interface between perception and cognition. Fodor, Churchland and Raftopoulos offer relatively impoverished accounts of perception in comparison with folk psychological accounts. Taking a different approach, Gallagher argues that perception involves more than brain processes and that by viewing perception as embodied, embedded, extended and enactive, it is "quite possible to develop an extremely rich and fruitful description of perception" (Gallagher: 2008a, 163).

Cognitive psychologists usually conceive of perception as being modular, at least in part, because there is a wealth of empirical data that suggests that low level sensory processes have modular characteristics. A module is a relatively autonomous structure which receives input from, and sends output to, other cognitive processes and structures. It is domain specific, genetically determined and associated with distinct neurological architecture. The apparent modularity of early perceptual processing raises a much debated issue, that is, whether or not low level perceptual modules are informationally encapsulated, or whether they are open to top-down influence by higher cognitive processes. The phenomenon of illusions is often cited as evidence supporting the claim that low level perceptual modules are informationally encapsulated because the misperceptions persist despite our understanding of the nature of the phenomena. If delusions are driven by perceptual anomalies that are encapsulated, it helps explain their persistence. However, the degree to which perception is penetrable by higher cognitive processes is still a matter of debate. Two influential protagonists in the debate are Jerry Fodor and Paul Churchland. A brief account of their debate will be given, followed by an explication of Raftopoulos' disagreement with Churchland's criticisms of Fodor, and an account of his own semi-Fodorian account. Finally, Shaun Gallagher's contrasting view will be discussed.

3. Jerry Fodor

Fodor (1983) proposes a structure of the mind consisting of two types of distinct architecture—an informationally encapsulated, innate, domain-specific, modular perceptual 'input-output' system, and a non-modular, domain-general, unencapsulated central system. According to Fodor, the encapsulation of the modules in the input system quarantines them from top-down influences from the central processing system and they are, therefore, theory-neutral.

> A module is (inter alia) an informationally encapsulated computational system – an inference-making mechanism whose access to background information is constrained by general features of cognitive architecture, hence relatively rigidly and relatively permanently constrained. One can conceptualize a module as a special-purpose computer with a proprietary database, under the conditions that: (a) the operations it performs have access only to the information in its database (together of course with specifications of currently

> impinging proximal stimulations), and (b) at least some information that is available to at least some cognitive processes is not available to the module. (Fodor: 2006, 518)

Fodor claims that the input systems are innately specified and have a fixed neural structure, and further, that the endogenous features of the human cognitive system, if not largely fixed at birth, are at least genetically pre-programmed and then triggered by the newborn's experience. The inability of the central processing system to feed information, top-down, to the modules helps support their rigidity, and thus enables the modules to produce more or less reliable responses to the environment. Responses need to be reliable because "perception is above all concerned with keeping track of the state of the organism's local spatiotemporal environment" and to immediately detect "what is available, for example, for eating or been eaten by" (Fodor: 2006, 519). According to Fodor, object identification is part of the fixed, or pre-programmed visual system. Through pairing objects with a 'form-concept' dictionary the visual system is able to deliver basic categorizations (Fodor: 1983). Fodor argues that the endogenously fixed, theory-neutral level of perception means all people share a common perceptual experience and that this facilitates communication. He most frequently cites the persistence of illusions in support of his claims because he thinks this argument is the most persuasive of the arguments for modularity and it is "certainly the shortest" (Fodor: 2006, 516).

> How much of what you know/believe/desire actually does affect the way you see? The persistence of illusions suggests the answer must be: "at most, less than all of it". (Fodor: 2006, 517)

He takes the Müller-Lyer illusion[38] as his example, pointing out that neither knowing the lines are of equal length, nor wanting the illusion to go away prevents one from continuing to see the lines as unequal. He claims this demonstrates that at least some perceptual processes are insensitive to some of one's higher cognitive processes. Fodor's theory is in contrast to domain general processing theories, in which mental activity is distributed across the brain and cannot be decomposed into independent units.

[38] The Müller-Lyer illusion consists of a set of lines that end in arrow heads or tails. Although the straight lines connecting the arrow head/tails are the same length, the different orientation of the heads and tails creates the illusion of the lines being of different lengths.

4. Paul Churchland

Churchland (1988) challenges Fodor's claims in two ways. He argues that even if the modules are encapsulated they are not necessarily theory-neutral. He also argues that the modules are not impenetrable. According to Fodor, there are features of cognition that resemble empirical assumptions about the world. Churchland argues that these assumptions, such as the three dimensionality of space, special and temporal continuity of common objects and the like, are not theory-neutral, and further, hard-wiring the assumptions does not make them so. He challenges Fodor's claim that modules are encapsulated in three ways. First he cites the example of 'inverted lenses'. When people wear lenses that invert the retinal image it only takes a week before the inverted image is reconfigured to cohere with data from the other senses, after which behaviour returns to normal. This example of the specific orientation of the visual field shows that perception can be penetrated by experience and the visual system can be reshaped by it. Churchland claims it demonstrates "the plasticity of some very deep 'assumptions' implicit in visual processing" (Churchland: 1988, 175). A second challenge is aimed at Fodor's argument from the Müller-Lyer illusion. Empirical data indicate that the Müller-Lyer illusion is the result of learning.[39] Whilst Churchland concedes the illusion cannot be resisted by a fleeting, voluntary attempt, he argues that if the illusion is susceptible to being learnt over a long period of time, in principle, it should be susceptible to being unlearnt over an extended period of time. A third argument Churchland presents is based on the discovery of 'descending pathways'. Neuropsychological evidence from cell staining not only identifies ascending neuronal pathways from peripheral modules to higher cognitive centres but also identifies descending pathways that lead, stepwise back through the

[39] "Findings provide evidence not only of variability across development with respect to the allegedly endogenous hypothesis of the visual input system responsible for our susceptibility to the Müller-Lyer illusion, but also of cultures in which many adults are not, in fact, susceptible to the illusion." (McCauley & Henrich: 2006, 5)

intermediate brain areas, perhaps even to the earliest processing systems at the retina.

> There are similar chains of descending pathways, from various areas topmost in the information-processing hierarchy, down through all of the intermediate processing stages and all the way out to the periphery, for all of the other sensory modalities as well. (Churchland: 1988, 177)

He suggests the pathways allow for the modulation of lower level neural activity as a function of the demands sent down from levels higher in the cognitive hierarchy.

Fodor (1988) resists Churchland's challenges to his impenetrability claims with varying degrees of plausibility. He points out that the type of plasticity demonstrated in the 'inverted lenses' phenomenon might be expected because, as we grow, we have to recalibrate the perceptual/motor mechanism that calibrates bodily gesture with perceived spatial position. Thus, he concedes to Churchland the plasticity of perception in specific ecological cases. However, he does not think this threatens his challenge because Churchland's penetrability requires that "you can somehow reshape the perceptual field by learning physics" (Fodor: 1988, 194). That is, *beliefs* must be able to reshape the perceptual field. On the subject of descending pathways he says "One thing is clear: if there is no cognitive penetration of perception, then at least "descending pathways" aren't for *that*" (Fodor: 1988, 194).

5. Athanassios Raftopoulos

Raftopoulos (2009) disagrees with both Fodor and Churchland. Unlike Churchland, he claims cognitive penetration does not permeate all levels of perception, and he criticises Churchland's criticism of Fodor. Although, like Fodor, he argues for an encapsulated level of perception, he cuts the interface between the encapsulated and unencapsulated systems in a different place. According to Raftopoulos, the earliest level of vision, the first forward sweep (FFS) and the locally recurrent processing (LRP) are encapsulated. He bases this on data provided by neuroscience regarding the timing of the various stages of visual processing. He argues from such data that the first 100-120 ms of

processing is immune from top-down influence by citing a wide range of experiments that show the top-down influences on perceptual processing do not impact on brain activity until after this time lapse. The output of early visual processing is fed forward and made available to cognition, but there is no feedback from higher levels of cognition to early visual processing that occurs in this time frame. Consequently, he suggests that the term 'perception' should be restricted to early nonconceptual visual processing and everything else be termed 'cognition'.

Raftopoulos argues for a modular organization of our perceptual systems which is somewhat different to Fodor. His account is based on "the distinction between noncognitive perception and cognitive observation" (Raftopoulos: 2001, 445). Like Fodor, he describes perception as domain-specific, fast, automatic, mandatory, independent of conscious control, and emerging as the result of the interaction between a constrained brain and the environment. However, he disagrees with Fodor's inclusion of basic categories in perception. According to Fodor, a 3D sketch is computed algorithmically in a bottom-up fashion, and an encapsulated processor performs object identification. Then, at the final stage, the 3D sketch is paired with a "form-concept" dictionary, which selects one of the outputs of the encapsulated processor. This contrasts with theories that claim the formation of object-centred representations relies on a flow of top-down information. Fodor distinguishes between 'fixation of appearance' or 'observation' (the result of the functioning of the perceptual modules) and 'fixation of belief' (the result of processing the output of higher cognitive modules). This is the distinction between what we see and how we interpret it. It is at this point that Raftopolous disagrees with Fodor. Raftopoulos restricts perception to non-cognitive processing, and Fodor's 'observation' (the output of the perceptual modules) includes cognition. Observation involves object recognition and object recognition is a cognitive process. Rather than being cut at the observation/cognition interface, Raftopoulos suggests the cut should be made at the perception/cognition interface.

Raftopoulos also disagrees with Churchland's argument against the impenetrability of the perceptual system based on an examination of various illusions and visual effects. In addition to the Müller-Lyer illusion already discussed, Churchland argues that the use of *context* to resolve the ambiguity in figures such as 'duck-rabbit', threatens Fodor's account. However, Fodor claims that the context does not determine the output of the modules but that the modules propose all possible syntactic analyses and higher-level processes select the most appropriate. All other analyses are deactivated and do not participate in further processing. This theory is filtering 'weak interactionism' and contrasts with Churchland's 'strong interactionism'. Raftopoulos supports Fodor in that he suggests a weak interactionist model can account for the results. He says the early processing produces a duck-rabbit configuration. About 70ms after stimulus onset, spatial attention, whether endogenously or exogenously driven, decomposes the figure such that the ensuing phenomenal content of perception is of a duck-like or a rabbit-like figure. This takes place in areas V4 and MT, in which shape is encoded. The content may eventually be conceptualised as either a duck or a rabbit and become available to awareness. This is compatible with Fodor's claim that whether one sees a duck or a rabbit depends on whether one is attending to the ears or the bill. Raftopoulos suggests weak interactionism can explain much, if not all of Churchland's evidence. Finally, although Raftopoulos agrees that the descending neural pathways track the ascending pathways all the way down, he stresses that the timing of the activation is important.

> The same areas that process sensory specific information in a bottom-up way are also involved in higher-level thought (voluntarily attention-driven search, imagery), except that in the latter case they are reentered in a top-down manner, and their activation is amplified. (Raftopoulos: 2009, 296)

Although top-down influence activates the same areas that are activated during the FFS and LRP, amplifying the recordings in those areas, the effects are delayed in time because the higher centres feed information from sensory systems back to the brain areas that processed the original signal. According to Raftopoulos, the time lag before the brain areas are re-activated preserves the impenetrability of early vision. Raftopoulos' account is the most impoverished

of the accounts of perception because in allowing no cognition, it allows no meaning content in his definition of perception.

The three authors discussed above all attempt to give an account of a perceptual system to which we have no direct access. They all attempt to describe the hidden processes that give rise to our conscious perceptual experience. But rather than hypothesising about the contents of the black box, as they do, with all the epistemological problems it involves, it may prove more fruitful to take our conscious perceptual experience as the starting point rather than the end point. Shaun Gallagher takes such an approach, and his account will be discussed in the following section.

6. Shaun Gallagher

The above three accounts of perception are impoverished in that they restrict the term 'perception' more-or-less to the sub-personal processing of sensory data. The degree to which this early-level processing might be influenced by higher level cognition is disputed, but the conceptual content of perception is limited in range from no content at all (Raftopoulos) to basic categories (Fodor). Gallagher defines perception in a fundamentally different way to the above writers. As a phenomenologist, his focus is on the subject's conscious experience of perceiving her world. Rather than attempting to divide the sub-personal processes into distinct levels, he begins with what we know, that is, with the output of all the sub-personal processes, both mechanical and cognitive, that produce our direct experience of perceiving our world. Gallagher's 'perception' has rich content.

According to Gallagher "we want to distinguish between sub-personal explanations of perception and conscious perception itself" (Gallagher: 2008b, 537). The distinction is necessary because "philosophers of mind and of cognitive science have tended to blur the distinction between content at the personal level and at the subpersonal level" (de Pinedo-Garcia & Noble: 2008, 87). But making the distinction is not the end of the matter because the two

levels can be defined in different ways. For example de Pinedo-Garcia and Noble see the distinction as "one between explanations that focus on an agent's interaction with its environment, and explanations that focus on the physical or computational enabling conditions of such an interaction"[40] (de Pinedo-Garcia & Noble: 2008, 87). Frankish makes the following distinction.

> Personal level states and events are ones that are properly attributed to a person or creature as a whole, rather than to some organ or subsystem.
> Subpersonal states and events are ones that are properly attributed to some organ or subsystem rather than to the person as a whole. (Frankish: 2009, 90)

Davies points to the explanatory gap between the levels, citing the most familiar—that associated with phenomenal consciousness.

> When we have been told all that there is to tell about the information processing going on in a conscious creature's brain this still seems to fall short of a satisfying answer to the question why there is something, rather than nothing at all, that it is like to be that creature. (Davies: 2000, 95)

Whilst acknowledging the complexity and importance of sub-personal processes, Gallagher focuses on the phenomenology of perception, the conscious experience of 'what it is like' to see something. Fodor, Churchland and Raftopolous offer sub-personal explanations of perception whereas Gallagher offers a personal level explanation of perception.

Gallagher's account is in keeping with Jaspers' view that "perceptions are never mechanical responses to stimuli; there is always at the same time a perception of meaning" (Jaspers: 1963, 99). In the phenomenological tradition 'meaning' is conceived of as affordance. "If I see a knife, I see a tool for cutting" (Jaspers: 1963, 99). "I see my car as drivable ... the driveability of my car is built into my direct perception of it" (Gallagher: 2008b, 537). Perception enriched with meaning is immediate or direct. Gallagher says of direct perception that despite the complexities of sub-personal processes and the ability of neuroscientists to 'carve up' the early visual system into processing areas, and whether or not top-down processes influence sensory information, when we consciously perceive, the perception is already informed by the relevant sub-personal processes. "My

[40] It is understandable that de Pinedo-Garcia & Noble make no mention of the phenomenal experience as their interest is in artificial life.

perception, in this sense, is direct even if the sub-personal sensory processing that underpins it follows a complex and dynamic route" (Gallagher: 2008b, 537). To flesh out this idea he contrasts two ways we might experience perception, naming one 'smart perception' and the other 'not-so-smart perception'. He distinguishes between the two by giving the following example of seeing his car.

> I open my eyes and see a certain unrecognized red mass with a specific shape just in front of me. My eyes are working fine, thank you. My visual cortex is processing all of the preliminary visual information, and what vision delivers is the meaningless red mass, which I then have to interpret in some non-visual, non-perceptual cognitive steps that go beyond perception itself. Let us call this not-so-smart perception. In contrast, in the very same situation, when I open my eyes I see my car. It is true that it has a specific shape and is red, and I do see the shape and the color, but I see the shape and the color as being aspects of something that is amazingly recognizable as my car. Actually, if you asked me what I see, I would likely not say that I see a red and shapely mass. Somehow I see through those aspects and I see my car. I do not see red mass, shape and color, and then try to piece all of that together to make it add up to my car. I simply and directly see my car. So let us call this a perception with some degree of smarts. (Gallagher: 2008b, 53)

Gallagher argues that perception is smart by appealing to our subjective experience of perceiving our world.

The richness of Gallagher's view of perception is far more than a liberal interpretation of the term 'perception', one that includes rich meaning as a component of the perceptual experience. Important to his account is the *direct nature* of the rich content and, as discussed in Chapter 1 of this thesis, the embodied, embedded and enactive nature of all cognition. Of further importance to Gallagher is the impact intersubjectivity has on the perception of objects. In fact he asks "is it possible to provide an adequate account of object perception without taking intersubjectivity into account?" (Gallagher: 2008a, 171). His argument is based on two issues, one involving development and the other the normal phenomenology of object perception. In support of the first, he points out that evidence from developmental psychology shows that we gain access to a meaningful world through our interaction with others. We start to learn about the world by seeing how others relate to objects in the world, and in this way learn to perceptually distinguish objects and to attend to certain objects in preference to others. The second argument relates to our awareness of occluded sides of objects. The original idea that an object is perceived at any given moment as possessing a plurality of co-existing profiles, attempted to account

for the fact that we bring our past knowledge of an object to our present perception of that object. For example, we know what a cup looks like from every angle because of our past experience of cups, and we bring that knowledge to our immediate perception of the cup in front of us. The problem with this explanation is that we do not see the present cup with respect to past profiles (ones we remember seeing), or to future profiles (ones we imagine we would see if we turned the cup around), for these lack the required actuality. That is, it is impossible for these profiles to be actualized in the *present*, and the plurality of profiles is conceived of as *co-existing.* It is not a combination of past, present and future.

> Husserl thus comes to an alternative account: if the absent profiles cannot be correlated with *my* possible but non-actualized perceptions, then the absent profiles may be correlated with the possible perceptions that others could currently have. (Gallagher: 2008a, 172)

According to Gallagher, co-existing profiles cannot be explained by remembering past experiences or imagining future experiences of viewing the object from different perspectives because this does not match the phenomenology of viewing an object. Rather, he accepts Husserl's theory of 'open intersubjectivity'. The theory basically proposes that "an analysis of object perception immanently refers us to the possible perception of a plurality of possible subjects" (Gallagher: 2008a, 172). Thus, it is not only the physical environment, but also our social environment that shapes the way we see objects. The issue of intersubjectivity expands the concept of meaning as affordance and, therefore, the concept of perception.

> I see something as something to be used, as others have used it or have failed to use it, and often in a context that includes others. I see it as meaningfully useful because it involves others, whether directly (as it may involve my co-workers) or indirectly (because certain cultural meanings and norms are part of its meaning). More generally, my perception of things and instruments, but also of contexts and places, and the world as such, is significantly invested with meanings and values derived from others. (Gallagher: 2008a, 176)

The above brief account does not do justice to the richness and complexity of Gallagher's view, but it is sufficient to draw a sharp contrast with the concepts of perception offered by Fodor, Churchland and Raftopoulos.

Gallagher's view of perception is much closer to the 'common sense' way in which the term is used than are the three impoverished account described

above. Used in the 'common sense' way, the term refers to the way one experiences or sees the world rather than to some low-level modular process that is inaccessible to consciousness. And the way we see our world is influenced by our culture, the immediate context, our intentions and desires, our physicality and so forth. It embraces *all* levels of sub-personal cognitive processing that are necessary to produce a full and rich direct perceptual experience. As Gallagher describes it, when I look at my car I do not see a red and shapely blob, I see my car with its driveability, and any personal meaning and emotional colouring that has become attached to it. In the same vein, when I look at my arm I see 'my arm' with all that this means. When I look in the mirror, I see 'my reflection' and all that that means. But in the case of delusions it seems that when the person looks at her arm she sees it as 'not my arm', even as 'someone else's arm', and when she looks in the mirror rather than seeing 'my reflection' she sees a stranger. Perception is clearly distorted in some way or at some level. What happens at the sub-personal or at the conscious level to cause the disturbed perceptual experiences is largely a matter of theory, hypothesis and speculation, all of which will be influenced by what one deems to be the limits of 'perception'. In this thesis I am arguing for an extremely rich view of perception, one that is fully informed by past experience, culture, context and intersubjectivity. I use the term 'perception' to mean the process by which a person attains awareness and understanding of her environment by organising and interpreting sensory information.

7. Perception, Meaning and Concepts

It is the interpreting of sensory data that gives the world meaning. Meaning refers to the sense of significance or value an object or event has to a person, and it is linked to psychological conditions such as wanting and intending, and to the conventions and rules germane to a person's culture. A full discussion of the theories of meaning is not within the scope of this thesis. As Miller notes "A full survey of the development of theories of meaning—in both the formal and informal senses—would require a full history of analytic philosophy itself" (Miller: 2011, 1). Meaning is drawn from the concepts we possess and, as the

development of specific concepts is pertinent to the argument I will make in following chapters, a clarification of the term 'concept' is necessary. As with 'perception' the term 'concept' means different things to different people. According to Machery, "by 'concept' philosophers refer to that which enables people to have a propositional attitude (beliefs, desires, etc) about the objects of their attitudes" (Machery: 2010, 603). In contrast, he describes four ways in which the notion of concept is characterised by cognitive scientists: as temporary bodies of knowledge in working memory, as bodies of knowledge under organismic control, as constituents of thought and as categorization devices (Machery: 2010, 602).

> Thus, although both philosophers and cognitive scientists use the term 'concept', they are not talking about the same things: cognitive scientists are talking about a certain kind of bodies of knowledge, while philosophers are talking about that which enables people to have propositional attitudes. (Machery: 2010, 604)

Having said that, the divide is not as clear cut as the above implies. As Laurence and Margolis stress, "in spite of these differences, there has been a significant amount of interdisciplinary interaction among theorists working on concepts" (Laurence & Margolis: 1999, 3). Nevertheless, many issues remain a matter of debate across disciplines. Rather than debating these issues, I will simply state the sense in which I intend to use the term. I use it in a cognitive science sense, that is, as temporary bodies of knowledge in working memory, rather than a philosophical one. More specifically I use the term to mean a general idea, usually mediated by a word, symbol or sign, and which may combine several elements from different sources into a single notion. I view the process of learning a concept as involving the abstracting of a quality or property from an object or event, and then generalizing that quality or property to all appropriate objects or events. In the following chapters, in keeping with Jaspers, I take delusion to be the manifestation of a disturbance in the meaning component of the immediate perceptual experience.

Whereas Jaspers presented his theory in relation to schizophrenia, the condition within which most delusions occur, I restrict my account to an exploration of the idea in relation to monothematic delusions. My example in this is the Capgras delusion.

PART TWO

"A compelling literature documents that there is much "physical" in "mental" disorders and much "mental" in "physical" disorders."

Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision
Washington, DC: American Psychiatric Association, 2000, pxxx

"It is something of as fantasy, to use Dennett's term, to suggest that neuroscience or psychology are best done by averaging out, reducing, and re-engineering first-person data so that it looks like third-person data."

Shaun Gallagher: 2003 Phenomenology and Experimental Design *Journal of Consciousness Studies, 10*(9-10) p98

INTRODUCTION

In this section of my thesis I offer a new account of both the first and second factor in a two-factor account of the Capgras delusion. I begin with a delayed-processing account of the first factor. The delayed-processing account is based on the Ellis (2007) face processing model. However, rather than locating the impairment that results in the Capgras delusion after face recognition, as Ellis does (see Chapter 6 Fig 1), I locate it upstream of face recognition, at the site of the initial structural encoding. I argue that a slowing of the encoding processing causes the face to be erroneously classified as unfamiliar. A conflict then arises when the target individual is identified as a particular person. I then argue that the conflict is resolved by the attribution of the rich 'impostor' concept. My account differs from other two-factor accounts of the Capgras delusion in that I suggest that the meaning component of the immediate perceptual experience is an interpretation of the face *in context.* It is not merely an interpretation of the face without reference to any other salient data. I argue that just as we can grasp the gist of the scene in milliseconds, we also can grasp the gist of the meaning of the scene with equal speed. It is this first pass interpretation of 'what is going on' that produces the impostor content.

A second difference between my account and other two-factor accounts of the Capgras delusion is that it is important how the 'impostor' concept that provides the meaning is formed. I argue that, for most people, the concept is formed through exposure to fiction. An additional reason to believe the Capgras patient might be viewing her loved-one from the perspective of fiction is that it helps explain some of the puzzling characteristics of the delusion. These are the incorrigibility of the delusion, its circumscribed nature, the phenomena of double-bookkeeping, the bizarre nature of the content, and the fact that some patients are passive whilst others are aggressive. The fictional nature of the Capgras impostor is basic to my account of the delusion. If the Capgras patient's impostor is, in fact, a fictional character, the nature of alternative realities and the way in which we engage in fiction become important issues. My discussion

of these issues includes Walton's (1990) description of make-believe and the construction of imaginary worlds, and Gallagher's (2009b) multiple realities hypothesis.

According to my account of the Capgras delusion, impaired face encoding creates a data conflict between familiar/unfamiliar. The conflict is resolved by the 'impostor' attribution, because both familiarity and unfamiliarity are implied by the concept. However, the way in which most people develop the concept suggests that the patient's impostor is a fictional character. As fictional characters do not live in the real world, the application of the (fictional) impostor concept to a person living in the real world creates a new conflict. It creates a cross-world conflict. As the conflict is created by an internal processing anomaly, it cannot be corrected by examining the environment more closely, such as we might do, for example, by measuring the lines in the Müller-Lyer illusion. Looking again only serves to reproduce the experience. A different strategy is needed to resolve the cross-world conflict, and I argue that we can learn about the second factor by understanding this strategy.

In my account, the second factor involves our normal ability to shift our perspective from one reality to another. Such ability enables us to enter all forms of alternative realities and imaginative states, including the alternative reality of fictional worlds. Empirical evidence from research of the hypnotic condition suggests that the mechanism that enables us to enter an alternative reality is the decoupling of a conflict monitoring system from executive control. My hypothesis is that when the Capgras patient is caught in a cross-world conflict created by the first factor, the cognitive dissonance it causes stimulates the automatic decoupling of the monitoring system and executive control. That is, a normal mechanism is brought to bear on an abnormal situation. When the decoupling occurs, the Capgras patient enters an alternative reality (a fictional world), which allows the application of the (fictional) impostor meaning to a real world person, to pass unchallenged. The decoupling takes place at the subpersonal level, enabling the (fictional) impostor concept to be presented to

consciousness as the meaning component of the immediate perceptual experience. Thus, the (fictional) impostor is experienced as being part of everyday reality. The patient 'sees' her loved-one as an impostor, rather than as him or herself, in much the same way as we 'see' the characters portrayed by actors in a play, rather than seeing the actors themselves. Normally engaging in fiction is a choice, and, therefore, under voluntary control. But for the delusional patient it is not a voluntary choice, and therefore, not under voluntary control. As the decoupling is stimulated automatically, the patient has no choice but to enter the alternative reality, and cannot choose to exit it. She becomes trapped in the fiction.

When engaging in fiction, the degree to which we become absorbed in the story determines the degree to which we experience the fictional characters and events as 'real'. The capacity to become absorbed in a story varies from person to person, and I suggest this reflects the extent to which each individual is able to decouple their monitoring system from executive control. I argue that the degree to which the delusion is developed and maintained is dependent on the degree to which the patient becomes absorbed in the fiction, and therefore, the degree to which she experiences the fiction as 'real'. I suggest that those who have a greater capacity to become absorbed in fiction are more likely to develop a delusion. This gives rise to the prediction that patients suffering delusions will score higher on the Tellegen Absorption Scale (TAS) than non-delusional subjects, and also, from those that have the first factor but do not develop the delusion.

The Absorption Scale (TAS)

The Absorption Scale (Tellegen & Atkinson: 1974) is one of the sections of the Multidimensional Personality Questionnaire. Absorption is a disposition or personality trait in which a person becomes absorbed in her mental imagery, particularly fantasy.

> Absorption is interpreted as a disposition for having episodes of "total" attention that fully engages one's representational (i.e., perceptual, enactive, imaginative, and ideational) resources. This kind of attentional functioning is believed to result in a heightened sense of the reality of the attentional object, imperviousness to distracting events, and as an altered

> sense of reality in general, including an empathically altered sense of self. (Tellegen & Atkinson: 1974, 268)

The ability to become absorbed varies from person to person. However, Tellegen and Atkinson suggest that experiencing the altered states associated with absorption could be "cultivated and elaborated into role-playing skills in the more specific interpersonal and theatrical sense" (Tellegen & Atkinson: 1974, 275). One of the purposes for developing the Absorption Scale was to enable Tellegen and Atkinson to test the putative link between absorption and hypnosis. They found their results supported Hilgard's hypothesis (E. Hilgard: 1965; J. Hilgard: 1970) that hypnosis involves 'imaginative involvement'. The Tellegen Absorption Scale is widely used to measure openness to absorbing and self-altering experiences.

CHAPTER 6: THE CAPGRAS DELUSION:
A delayed-processing account of factor 1

1. Introduction

In this chapter I offer a new account of the first factor in a two-factor account of the Capgras delusion. The Capgras delusion involves the belief that familiar people in the patient's life, and in rare cases pets or inanimate objects, have been replaced by identical doubles. The generally accepted hypothesis is that the delusion results when the neuro-anatomical pathways responsible for producing an appropriate autonomic/affective response to familiar stimuli become disconnected from the face processing system. The hypothesised impairment is supported by skin conductance (SCR) data that shows Capgras patients as a group record a lower than normal response to familiar faces. Coltheart goes so far as to say "it seems difficult to deny that the failure of autonomic response to familiar faces that occurs in the Capgras delusion plays a causal role in the genesis of the delusion" (Coltheart Menzies & Sutton: 2010; 266). I challenge this point of view and offer an account of the Capgras delusion in which the autonomic/affective system is not impaired, nor does it fail. According to my account, impaired encoding of sensory data results in a person (or object) being erroneously classified as unknown[41]. The reduced autonomic/affective response is the normal response to such classification. The putative impairment I describe is drawn from the Whittlesea and Williams (2000) discrepancy-attribution theory of the source of the subjective feeling of familiarity.[42]

[41] There is evidence that most, but not all, Capgras patients have face processing difficulties. It is a slowing of the speed of structural encoding of a loved-one's face that I suggest causes the face to be misclassification as 'unfamiliar'.

[42] For the remainder of the chapter, in keeping with Coltheart, I will use 'autonomic' response when referring to the sub-personal process and 'affective' response when referring to conscious feelings.

2. The Capgras Delusion

The Capgras delusion is usually defined as the belief that a significant person in the patient's life, frequently a spouse or close relative, has been replaced by a visually similar impostor. The impairment occurs most commonly during visual processing, although there are rare cases of auditory Capgras in both blind (Reid, Young & Hellawell: 1993; Hermanowicz: 2002) and sighted (Lewis et al.: 2001) patients, and haptic Capgras, cases in which patients claim their loved one does not feel (physically) the same when touched (Rojo et al.: 1991; Ellis et al.: 1996). It was first described in detail by Capgras and Reboul-Lachaux in 1923. Although predominantely associated with neurodegenerative disease, the delusion is also associated with cerebrovascular disease, head trauma, epilepsy, toxicity and psychiatric disorder. Bauer's (1984) seminal work on prosopagnosia underlies the generally accepted hypothesis that the Capgras delusion results from a disruption to an aspect of the face recognition process. Prosopagnosia is a modality-specific disorder of recognition that is not due to sensory dysfunction, unfamiliarity with the stimulus or aphasic misnaming. It results in the patient's inability to recognise familiar faces. Using skin conductance response (SCR) tests, Bauer discovered that some prosopagnosic patients exhibit covert recognition of familiar faces in the absence of the overt identification of those faces. Bauer's work prompted Ellis and Young (1990) to propose that the Capgras delusion might be the mirror image of prosopagnosia. This would mean Capgras patients receive a veridical image of the person they are looking at, which stimulates the appropriate semantic data, but they lack the information that Bauer suggests may carry some sort of affective tone. The face is recognised as belonging to a known person (looks familiar) but there is no affective data (doesn't feel familiar). The delusion reflects the patient's attempt to makes sense of the fact the stimuli no longer have appropriate affective significance. The clear prediction that follows is that Capgras patients will not show the normal SCR to familiar faces, despite the fact that the faces will be overtly recognised (Ellis & Young: 1990). The prediction was confirmed (Ellis et al.: 1997; Hirstein & Ramachandran: 1997). In SCR tests, Capgras patients did

not reveal the normal autonomic discrimination between familiar and unfamiliar faces. However, magnitude and rate of initial habituation, and orienting responses to auditory tones were normal, showing that hyporesponsiveness to familiar faces is circumscribed.

Following the work of Ellis and Lewis *et al.* (2000) and Breen, Caine and Coltheart (2000), Ellis (2007) revised the existing models of face recognition to accommodate the putative role affective response plays in normal face recognition and in the Capgras delusion. It can be seen from this model (see Fig.1) how breakdown in specific functions within the processing system might lead to both prosopagnosia and the Capgras delusion. By dividing the face recognition unit into two units—personal identity nodes and affective response to familiar stimuli—it allows for the possibility that a lesion disrupting the function A could lead to prosopagnosia whilst a lesion disrupting B could lead to the Capgras delusion.[43] It also shows how a disruption at C might cause a response at B to fail to produce a skin conductance response.[44]

[43] Robyn Langdon points out that the model can also be used to explain different types of prosopagnosia (personal communication).

[44] In Coltheart's (2010) account, the conscious registering of a lack of affective response is eliminated. Reduced affective response is replaced by reduced autonomic response. As people are not aware of their autonomic responses, conscious awareness of the anomaly does not play a causal role in the formation of the delusion as it does in the Ellis and Young (1990) account. "In many forms of delusion, something abnormal occurs of which the person is not conscious, unconscious processes of abductive inference are invoked to seek a hypothesis which, if true, would explain that abnormality, and a hypothesis is found which is judged satisfactory by these unconscious inferential processes. After all of that unconscious processing has been completed, the hypothesis is accepted as a (delusional) belief, and enters consciousness" (Coltheart, Menzies & Sutton: 2010, 264-265). Despite this change, the Ellis (2007) diagram can still be used to represent Coltheart's account if 'affective' response to familiar stimuli is changed to 'autonomic' response to familiar stimuli. Thus, a disruption at B would result in reduced autonomic response to a familiar face, and the attribution process would take place at the subconscious level rather than at the conscious level, as it does in the Ellis and Young (1990) account.

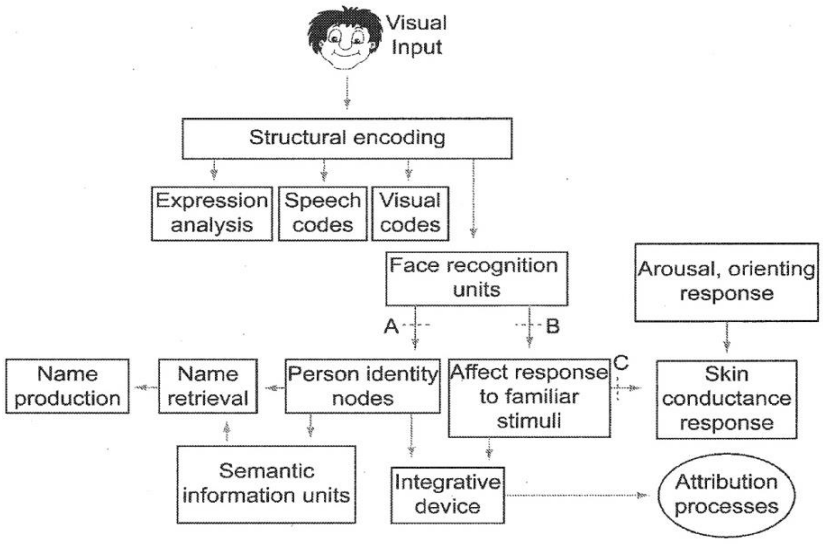

Figure 1: The Ellis face processing model shows how a lesion at A could result in prosopagnosia, and a lesion at B result in the Capgras delusion.

Although the above model is compelling, there are some characteristics of the delusion, and anomalies in the data on which it is based, that require further explanation. I will address these later in this chapter when I will argue that the account I am presenting provides a more plausible explanation of these issues.

3. The Delayed-Processing Account

i) Whittlesea and Williams

The delayed-processing account I propose, is an account of factor one in a two-factor account of the Capgras delusion. It is drawn from Whittlesea and William's (2000) discrepancy-attribution theory of the source of the subjective feeling of familiarity. It is important to note at the outset that Whittlesea and William's theory purports to explain the *conscious* feeling of familiarity, that is, the subjective feeling of having encountered a stimulus previously. They do not give an account of the Capgras delusion. According to their account, the fluency of the subpersonal processing of any given sense data is monitored. If the fluency meets expectations there is no conscious feeling of familiarity, but if there is a violation of expectation caused by a delay between processing the sensory input and accessing the accompanying details, it triggers a conscious

awareness of familiarity. My account is drawn from their description of how this theory specifically applies to face-processing.

According to Whittlesea and Williams, in extended acts of processing, people integrate the informational contents of various parts of their experience into larger organizations, producing more elaborate cognition. At the same time they evaluate the quality and coherence of their processing, which enables them to detect any errors or inconsistencies in that processing (Whittlesea & Williams: 2001). They propose that in evaluating their processing, people arrive at one of three major conclusions: that everything went well (coherence), that something went wrong (incongruity) or that something strange happened (discrepancy) (Whittlesea & Williams: 2000).

To make judgements based on the fluency of processing implies that people are able to make a comparison between their current performance and some norm or standard. In principle there are two standards, a general and a specific, that may be used. The ease of processing an item could be compared with the ease of processing other stimuli of the same type or class (general), or with the expectation that the person develops about how efficiently they should be able to process a particular stimulus (specific) (Whittlesea & Leboe: 2003). There are also two ways in which the fluency can be assessed. One is "the *speed* of processing", that is "the absolute magnitude of fluency [that] has direct impact on the decision process", and the other "the *discrepancy* between what is expected and what happens in a processing event" (Whittlesea & Leboe: 2003, 63; Whittlesea & Williams: 2000, 548). The Whittlesea and William's discrepancy-attribution theory of the source of the subjective feeling of familiarity is based on the following hypothesis.

> When the various components of a processing experience are perceived to fit together coherently, no particular subjective state occurs: People simply continue with their current activity. However, when a discrepancy is perceived between two parts of an experience, people switch over to an introspective, reflective state, attempting to reconcile the discrepancy and attribute it to some causal factor. In doing so, they experience a phenomenological reaction. (Whittlesea & Williams: 2000, 562)

According to Whittlesea and Williams (2001), it is the *fluency* of processing that is the relevant quality evaluated when an inference of familiarity is being made and this relates to face-processing in the following way.

> People expect known faces to be fluently processed and nonfluently processed faces to be unknown; that is, they expect a match between the fluency of perceiving and coming-to-mind of identity information. (Whittlesea & Williams: 2000, 548)

But the production of the *feeling* of familiarity it is not a simple case of 'fluent-therefore- familiar'. Studies by Jacoby and Whitehouse (1989) suggest that when people are aware of a reason why processing might be especially fluent, for example, when they are aware of priming, they do not experience a feeling of familiarity. Whittlesea and Williams (2000) argue that it is only when the fluency of processing *violates expectations* that we are consciously aware of the feeling of familiarity.

> In making recognition decisions, people appear to be impressed not by the fluency of processing per se but by the difference between their actual fluency and the fluency that could ordinarily be expected for that item in that context. (Whittlesea & Williams: 2000, 548)

Put simply, to experience a sense of familiarity, something needs to go wrong. Consequently, "The feeling of familiarity as a strong, conscious subjective state is a fairly rare event in ordinary life" (Whittlesea & Williams: 2000, 560).

In the Whittlesea and Williams' account the feeling of familiarity is stimulated when there is a delay between the processing of the visual input and the accessing of the person's details or personal identity nodes (PIN). Explaining this in terms of the Ellis face-processing model is to say that if there is a processing delay at A (but not at B) the autonomic response to the familiar face that results from B, stimulates a phenomenal experience of familiarity. The more familiar the face, the stronger will be the subjective feeling of having encountered the face before. This gives rise to the common experience of recognizing that one knows a person because they feel familiar, but being unable to recall who they are. Leboe and Whittlesea (2002) note that the process of recall is often initiated by a spontaneous experience of familiarity, and then proceeds intentionally and deliberately.

ii) The Capgras delusion

According to the Ellis and Young account, Capgras patients appear to have the opposite problem to the one described above. Rather than feeling the person is *familiar* but being *unable* to recall their identity, the Capgras patient feels the person is *unfamiliar* but is *able* to recall their identity[45]. One way to account for this difference is to suggest that the delay in processing occurs at a different stage in the processing. In the delayed-processing account I offer, the impairment in face-processing in Capgras patients occurs much earlier in the process than in the Ellis and Young and Whittlesea and Williams' models. In both these accounts the problem occurs *after* face recognition. My suggestion is that the problem occurs during the structural encoding of the raw face data, that is, *prior* to face recognition and the subsequent accessing of the personal identity nodes (PIN).

According to the delayed-processing account, past encoding of a specific face sets a standard against which the speed and fluency of present encoding are matched. In keeping with Whittlesea and Williams, I suggest that a comparator monitors the speed of the structural encoding and classifies a face as 'familiar' or 'unfamiliar' based on the principle that a familiar face will be rapidly encoded and an unfamiliar face will be more slowly encoded. It sets up a *prima facie* expectation that the person is known or unknown. There is an immediate autonomic response to this initial classification that is reflected in the SCR. The encoding speed will vary according to the degree of familiarity, with the most familiar face lying at one end of a continuum and an unfamiliar face lying at the other.

[45] Bayne and Pacherie (2004, 4) suggest that the Capgras patient does not merely suffer a loss of affective response but experiences a "disturbing feeling of unfamiliarity and estrangement". For a discussion of the feeling of unfamiliarity in the Capgras patient see Ratcliffe (2007).

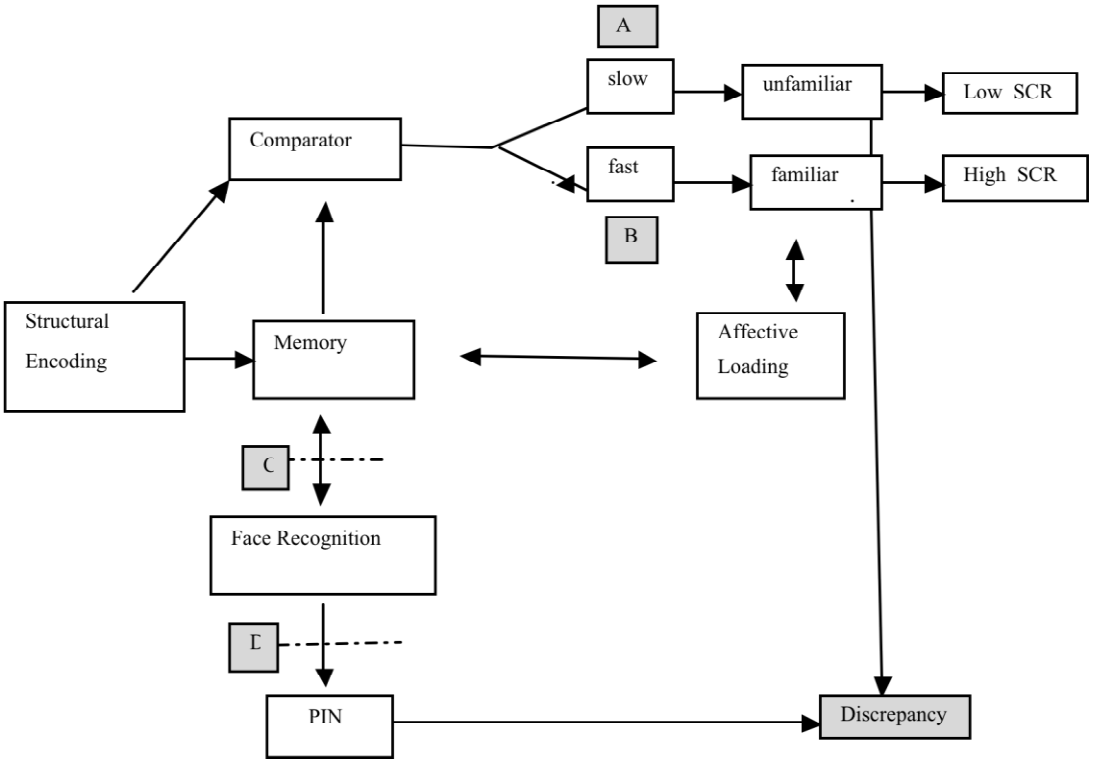

Figure 2: A and B represent the opposing ends of a continuum. C and D represent the sites of impairment with C leading to prosopagnosia and D to the feeling of familiarity described by Whittlesea and Williams. The discrepancy is between the classification of the face as 'unfamiliar' and the accessing of the PIN.

In the Capgras delusion, impairment that slows the speed of the structural encoding causes the comparator to erroneously classify a known face as 'unfamiliar', setting up the *prima facie* expectation that the known person is a stranger. However, although the process is delayed, it *does* produce a face. As the face is, in fact, known, the recognition unit is able to find a memory match for it, leading to accessing of the PIN. The slow processing causes the face to be erroneously classified as unfamiliar, but the memory matching of the face and the accessing of the PIN produces an identifiable known face. Thus, a conflict is created between the two sources of information, that the face is known and that it is not known. Figure 2 shows how it is possible for a slowing of the structural encoding to cause a known face to be classified as unfamiliar (A) without impacting on the process of accessing the PIN, and consequently, how this creates a discrepancy. It also shows how the slowing of the structural encoding of a face could lead to a delusion without it being necessary for the

autonomic/affective system to fail. Although aberrant autonomic response is implicated in the formation of the delusion, the system does not fail or become disconnected from the face processing system. It responds as it should respond when a face is classified as unknown.

The discrepancy created when a known face is classified as unfamiliar needs explaining.

> As practiced users of our memory systems, we have specific, often context dependent implicit expectations of the ease of memory processing – processing fluency – that may be violated in certain circumstances. Individuals are motivated to seek the most natural or salient explanation for unexpectedly more (or less fluent) production. (Barnier, Dienes & Mitchell: 2008, 157)

According to Whittlesea and Williams, a discrepancy between two parts of an experience causes a switch to an introspective, reflective state and the search for an attributional causal factor (Whittlesea & Williams: 2000). This fits quite nicely with an explanationist account of the Capgras delusion based on the delayed-processing model. The slow speed of the structural encoding sets up a *prima facie* expectation that the face is unknown. This expectation is violated when the identity details are accessed. The violation stimulates a phenomenal reaction, (subjective feeling of unfamiliarity), with a switch to the reflective mode and a search for an explanation as to why the person looks familiar but feels unfamiliar. However, the explanationist account is not without problems. When we believe we know someone, but cannot remember her identity, we consider the feeling of familiarity we experience sufficient justification for holding the belief that we know her. If, as Bayne and Pacherie (2004) suggest, Capgras patients experience a disturbing feeling of unfamiliarity, we might assume they would consider this sufficient justification for holding the belief that their loved-one is a stranger. But Capgras patients are typically at a loss to explain why they hold that belief. And as Pacherie (2009) notes, when pushed, the reasons they offer relate to minor differences in appearance or behaviour, not to their feeling of unfamiliarity. According to Whittlesea and Williams' theory, the stimulation of the phenomenal response (feeling of familiarity) is for the purpose of engaging conscious reflection in an attempt to solve the problem (forgotten identity). Likewise we could say that a phenomenal response (feeling

of unfamiliarity) in the Capgras patient would be for the purpose of engaging conscious reflection in an attempt to resolve the stranger/loved-one problem. This is indeed what explanationists claim. The person looks familiar but feels unfamiliar, and the post reflective judgment that the person is an impostor resolves that anomaly. But the fact that patients neglect to mention the feeling of unfamiliarity when accounting for their belief suggests that they do not have this phenomenal response. In the next chapter I will argue that the known/unknown conflict is resolved at the *sub-personal* level when the rich 'impostor' attribution is made. If this were the case, there would be no need for a phenomenal response because there would be no need for post-reflective problem solving. The fact that patients appear to lack the feeling of unfamiliarity supports this view.

4. Accounting for the SCR Data

Any account of the Capgras delusion has to accommodate the SCR data. That is, it needs to be able to account for the fact that, although Capgras patients as a group record a lower than normal SCR to all familiar faces, not all familiar faces are misidentified. For example, Capgras patients may misidentify a close relative but not doctors and nurses who are also familiar to them. The generally accepted hypothesis is that it is only the absence of the normally strong response to very familiar faces that triggers the delusion.

> We suspect that the patients initially only voice concern about those cases where the discrepancy is perceived to be the greatest, which will usually be for the closest relatives. (Young: 1992, 54)

The absence of what would normally be a strong affective/autonomic response to a spouse, for example, is more noticeable than the absence of the normally weaker response to a familiar person who is not emotionally close to the patient.[46] The delayed-processing account offers a different first factor, and,

[46] According to Ratcliffe (2007; 2010), a Husserlian 'horizon of possibilities' with emotional tone surrounds objects and people, giving rise to a background of (inconspicuous) feelings that are normally not in conscious awareness. Rather, it is the absence of the feeling that is noticed. I argue that the presence or absence of affect, whether conscious or not, is not *causal* in the Capgras delusion. As this includes 'inconspicuous feelings' their presence or absence would not alter my argument.

therefore, a different explanation is required. What needs to be considered here is that many factors contribute to rapid encoding, including multimodal information and cues such as context, gait, gesture and the like. When the probability is high that a certain context surrounds a visual object, the processing of that object is facilitated. Objects are recognised more accurately (Davenport & Potter: 2004) and *faster* (Ganis & Kutas: 2003). It is acknowledged that recognition of other sensory information such as gait, or voice, feeds into the face recognition system at some point (Young, Ellis & Szulecka: 1990). There is considerable evidence Capgras patients have face-processing impairment.[47] If such impairment causes difficulty in the encoding of the raw face data, one can argue that skill in utilising cues could enable patients to compensate for this deficit. Skilled use of cues could enable all but the most familiar faces to be processed fast enough to be classified correctly. In the laboratory setting, in which photographs of familiar people are used, most cues to face encoding are eliminated by masking. This means patients are forced to rely on their impoverished structural encoding skills to produce a face. What I suggest is that in the laboratory, the speed of encoding familiar faces would be slowed sufficiently to cause all, or most, familiar faces to be erroneously classified as 'unfamiliar', with the SCR reflecting that classification (see Figure 3C). This slowing of face-processing would not occur in the everyday situation, thus giving rise to the prediction that in the everyday situation most familiar faces are not erroneously classified. This is because, in the everyday world, using cues may increase the encoding speed sufficiently to enable the correct classification of a less familiar face, but not sufficiently to create the high speed necessary to correctly classify a highly familiar face. It is a testable prediction because SCR readings could be taken while the patient is viewing the loved-one who is the target of the delusion (the real person), and while they are viewing, for example, a member of the hospital staff (the real person) whom the patient recognizes, but would not consider to be very well known to them. The prediction is that in response to the loved-one, the patient would record a

[47] Quoting twenty-two studies, Edelstyn and Oyebode (1999) note neuropsychological investigations of face-processing skills in Capgras patients have consistently reported the presence of deficiencies.

reduced SCR (stranger), but would record a normal SCR to a familiar face when viewing the staff member.

5. The Tendency of the Delusion to Spread Over Time

A characteristic of the Capgras delusion is that it tends to spread to other relatives and less familiar people over time (Capgras & Reboul-Lachaux: 1923; Todd *et al.*: 1981; Edelstyn & Oyebode: 1999; Young: 2000; Pacherie: 2009).

> The Capgras delusion typically is initially about one close relative, but as time passes, tends to spread to other relatives. The subject starts by thinking that his wife has been replaced and ends up thinking that his whole family has been replaced. (Pacherie: 2009, 121)

The tendency of the delusion to spread over time is not confined to the misidentification of family members. Mme M. described by Capgras and Reboul-Lachaux, first believed her daughter had been replaced, but over time "this delusion spread to the rest of the people in her circle" (Capgras & Reboul-Lachaux: 1923, 127). This included her husband, the concierge and all other tenants in her building, the domestic staff, nearly everyone at the hospital (doctors, nurses, and patients), the Prefect of Police and the commissariat. In another more striking case the misidentification spread, over time, to the inhabitants of nearly the entire town (Ellis *et al.*: 1997).

If the *strength* of affective response is used to explain why not all familiar people are misidentified, it makes it difficult to explain the tendency of the delusion to spread to those of less emotional significance. One might suggest that having experienced one misidentification (impostor), the patient becomes more attentive to the discrepancy between usual and present affective responses and, as a consequence, misidentifies additional people. However this explanation becomes increasingly implausible as the delusion spreads to people of little to no emotional significance, such as the inhabitants of nearly the entire town or nearly everyone at a hospital. I suggest the delayed-processing account offers a more plausible explanation of the delusion's tendency to spread over time. If we assume the structural encoding process is continuum based – the more familiar, the more fluent the processing – the degree of impairment might differentially affect degrees of familiarity.

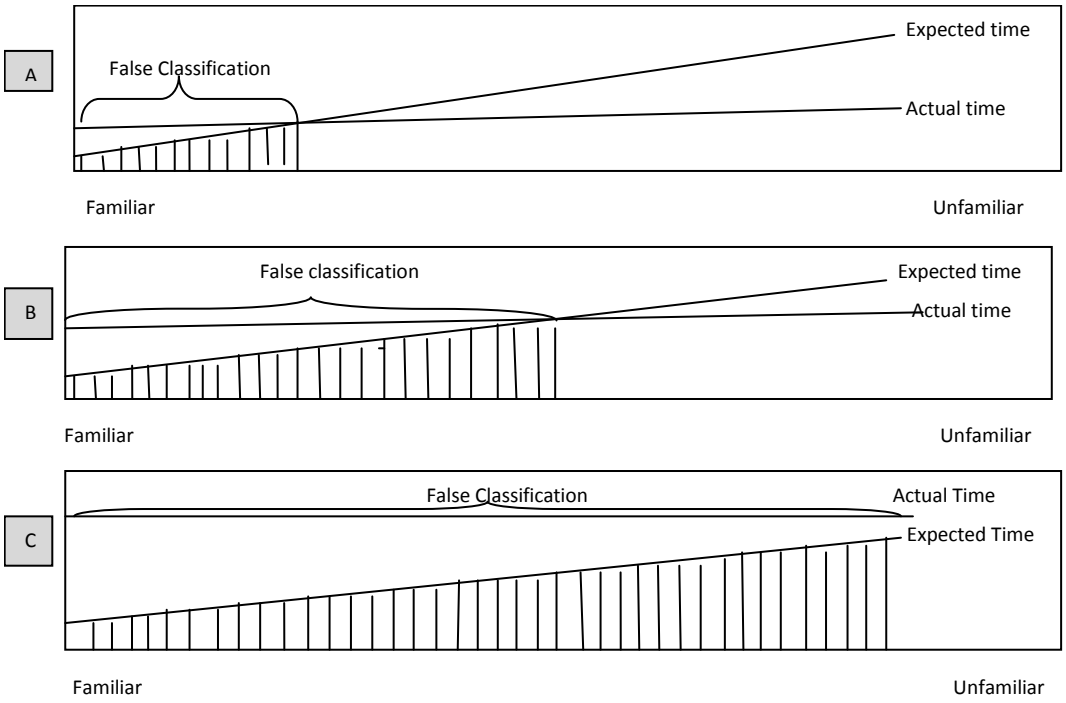

Figure 3: (A) Structural encoding slower than normally would be expected for a specific face causes the face to be falsely classified as 'unfamiliar'. (B) Deterioration of the condition (increased reduction in speed) causes less familiar faces to be falsely classified. (C) Extreme condition artificially produced in the laboratory, in which all, or most, familiar faces are falsely classified.

Because of repetition priming and context, we would expect the structural encoding of the face of a spouse, child or parent to be more fluent than that of other familiar faces. Arguably, more fluent processes would be more sensitive to a slight impairment. But if the condition deteriorates over time, the increased impairment would begin to impact on the normally less fluent processing of the faces of less familiar people (see Figure 3). The Capgras patient could begin with the misidentification of a spouse or other close relative but, if the condition deteriorates, misidentify an increasingly wider population over time. In its most extreme condition the erroneous classification of known people could extend to anyone familiar to the patient. If a patient has lived her life in a small town, most of the inhabitants, even if not well known, would at least be familiar to her and therefore, susceptible to misidentification as the condition deteriorated. It also could be argued that all of those misidentified by Mme M. were familiar to her and likewise vulnerable to misidentification due to an advanced condition.

6. Delayed-processing and the Erroneous Classification

There is a question specific to the delayed-processing account that needs to be answered, that is, why it is that face recognition does not trigger an affective loading that overrides the initial erroneous judgement. According to the model presented, it is because there is no feedback mechanism that enables this to occur. Fast modular responses evolved as survival responses. They enable an organism to rapidly respond to events in the environment. Such responses operate on the Hebbian principle summarised as 'what fires together wires together'. According to Hebb, "any two cells or systems of cells that are repeatedly active at the same time will tend to become 'associated', so that activity in one facilitates activity in the other" (Hebb: 1949, 70).

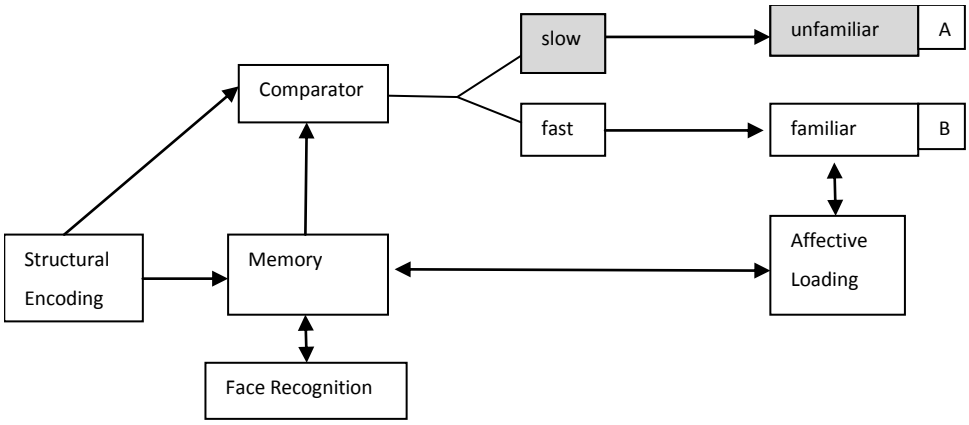

Figure 4: A. An 'unfamiliar classification is not followed by recognition therefore, no feedback loop develops. B. Processes that repeatedly occur together enable feedback loops to develop.

When a face is classified as familiar it is normally followed by face recognition, identity details and an accompanying affective loading. The repeated 'firing' sequence of the processes creates an association between them that enables a feedback loop between 'familiarity' and the other details to evolve (see Figure 4.) Information can travel in either direction in a feedback system. Thus, if information travels in one direction, a classification of 'familiar' with its accompanying affective loading can stimulate memory to produce an identity.

When it travels in the other, face recognition can stimulate the memory of affective loading, thus confirming that the face is familiar. In contrast, when the judgment 'unfamiliar' is made, there are no accompanying details of identity or affect stored in memory, and therefore, no firing sequence involving these can follow from a classification of 'unfamiliar'. Consequently, a feedback loop into the identity/affect memory system cannot evolve. Once the face has been classified as unfamiliar, the processing stream remains isolated from any memory of identity and affective loading. In the delayed-processing account of the Capgras delusion, it is hypothesised that the reduced speed of structural encoding causes the face to be erroneously classified as unfamiliar. This means there is no feedback loop into the memory system. As can be seen in Fig 4, even if the face is eventually recognised, the accompanying affective loading cannot impact on the initial classification, and therefore, there is no mechanism for overriding the erroneous 'unfamiliar' judgment at this stage of the process.

7. Capgras for Inanimate Objects

The hypothesis that the strength of the affective/autonomic response is a significant factor in the formation of the Capgras delusions is challenged by the rare cases of Capgras for inanimate objects because many of the misidentified objects are of low emotional impact. The original Ellis and Young hypothesis included all sense modalities (Ellis & Young: 1990), but the fact they specifically used the visual processing of the face to argue their case appears to have led to a split in the literature. Some researchers predominately define the Capgras delusion as being specific to people (Bourget & Whitehurst: 2004; Brighetti *et al.* 2007; Coltheart: 2007; Dietl, Brunner & Friess: 2003; Doran: 1990; Josephs: 2007; Tamam *et al.*: 2003; Young, G. : 2008), whilst others extend the syndrome to include inanimate objects such as household furniture (Ellis *et al.*: 1996; Nejad & Toofani: 2006) and animals, particularly pets (Ellis *et al.*: 1996; Rosler *et al.*: 2001).

> Although it is repeatedly claimed that the Capgras delusion is specific for people, particularly those emotionally closest to the patient, the literature clearly shows this is not the case. (Anderson: 1988, 696)

This is a moot point. Clearly, the hypothesis that the Capgras delusion is caused by a disruption to, or failure of, the affective/autonomic system during *face-processing* may be applied to people and animals, but not to inanimate objects. Consequently some researchers regard the delusional belief that inanimate objects have been replaced by near-identical substitutes as a separate delusion rather than a variant of Capgras. On the other hand, the similarities in the manifestation of the delusion suggest that it should be included under the Capgras umbrella. This remains an unresolved issue in the literature.

Classifying the misidentification of inanimate objects as a Capgras variant is consistent with Ellis and Young's original hypothesis that the delusion reflects the patient's attempt to make sense of the fact the stimuli no longer have appropriate affective significance. But because it was thought necessary to hypothesise that it is only the absence of *strong* affective responses that trigger the delusion, explaining Capgras for objects on the Ellis and Young model became more difficult. Capgras for pets can be included in the familiarity/emotional attachment hypothesis because pets have faces that can be recognised and are often the objects of deep emotional attachment. However, strictly speaking, Capgras for pets does not fit the Ellis and Young model because there are differences between processing a human face and that, for example, of a canary.[48]

Although it is clearly possible for patients to have a strong affective response to well loved possessions, it is not clear if the response to all the objects misidentified would be strong enough to trigger the delusion if absent. Anderson (1988) describes a case in which a man believes a wide range of household and personal object, including screws and paintbrushes, are misidentified. It is doubtful the man held great affection for these objects. Some researchers claim it is significant that in this and several other cases, the replacement objects are described as being older or more worn than the original. The observation gave rise to the hypothesis that the misidentification

[48] A case of Canary Capgras was reported by Rosler *et al* (2001).

of inanimate objects is caused by a failure to update memories and therefore it is not a Capgras variant. But the problem with the hypothesis is "it fails easily to account for those minority instances in the literature where the duplicate objects seem newer and better" (Ellis *et al.*: 1996; 36). As well as some object being described as newer or better, there are a range of other differences described by patients, such as colour, weight, pattern and design, that do not relate to age (Ellis *et al.*: 1996). The question to be considered here is whether the patients are describing an actual perceptual experience (*looks* different), or whether they are confabulating to explain an anomalous sense (*seems* different). If they are describing a perceptual experience, not all cases can be explained as a failure to update memory because not all objects look older. If the explanations provided by patients are confabulations designed to account for loss of a normally strong affective response, it remains to be explained how this could be the case for such items as microwave ovens, refrigerators and the like, perhaps even screws and paintbrushes. One way to address this problem is to suggest that the sensitivity to missing affective response might be related to the particular system. There might be a difference between the *object* recognition system and the *face* recognition system such that objects of low emotional impact could trigger the delusion if the affective response were to be reduced. However, other than serving to explain delusions for inanimate objects, there is no reason to suppose that this is the case. The delayed-processing account avoids the above problems because the strength of the affective/autonomic response is not relevant. As it is the reduced speed of the structural encoding of raw sense data that is causally relevant, the salient factor is not how one feels about an object but how often and closely one has viewed it, and therefore, how quickly it can be encoded.[49] If the relevant factor is speed of processing, rather than strength of affect, the same explanatory account can be given for all variants of the Capgras delusion.

[49] Some objects need to be more closely viewed to use than others (e. g. microwave compared to chair). Strong affection for an object also may be cause for both repeated and close viewing.

8. Summary and Conclusion

Bauer's investigation of prosopagnosia inspired Ellis and Young to formulate a hypothesis that the Capgras delusion might be caused by the reduction in the normal affective response to familiar stimuli, particularly to the face. The hypothesis appears to be supported by SCR data that shows Capgras patients register lower than normal responses to photographs of familiar faces. To explain why, in some cases, all familiar faces produce a reduced SCR but not all lead to a delusion, it was further hypothesised that it is only the absence of a normally strong affective response that triggers the delusion. However, this solution created another problem because it then became difficult to explain the characteristic tendency of the delusion to spread over time to less familiar people, in some cases, even to those to whom the normal affective response would be minimal. In addition, the hypothesis makes it more difficult to explain Capgras for inanimate objects, particularly for objects of low emotional impact. As a solution I offer a delayed-processing account of the first factor in a two-factor account of the delusion. The model draws on Whittlesea and Williams' discrepancy-attribution theory of the source of the subjective feeling of familiarity, more specifically, on their suggestion that in extended acts of processing, the speed and fluency of processing of an item is compared with the speed and fluency of past processing of that item, or class of items. Any violation of expectation causes the process to be classified as discrepant. I propose that a reduction in the speed of processing raw sensory data, usually from a face, causes the Capgras patient to erroneously classify a person (or object) as 'unfamiliar' whilst not impacting on the accessing of the identity details. This creates a known/unknown data conflict. In keeping with Whittlesea and Williams' theory, when a violation of expectancy is registered, a search for a resolution of the discrepancy is instigated. However, unlike the Whittlesea and Williams' account, in which a phenomenal response is generated to engage conscious reflection, I argue that patients' failure to offer the feeling of unfamiliarity as justification for their belief gives us reason to believe the discrepancy in the case of the Capgras delusion is resolved at the unconscious

level. In the delayed-processing account, the tendency of the delusion to spread over time is accounted for by deterioration in the condition in which reduced speed of structural encoding causes less familiar faces increasingly to be erroneously classified. The misidentification of inanimate objects is easier to include as a Capgras variant in the delayed-processing account, because the salient factor is the patient's familiarity with the object, not her affection for it. Patients could for example, rapidly encode the visual data from their microwave oven without necessarily holding any fondness for it, and a slowing of the encoding could lead to misidentification. The delayed-processing account offers a new way of explaining how the conflicting data (known/unknown) occurs in Capgras patients as the result of a face processing impairment. The next step in the process, the attribution of meaning, will be discussed in the following chapter. I will argue that the conflict is resolved at the sub-personal level when a rich 'impostor' attribution is made.

CHAPTER 7: MEANING AND THE CAPGRAS DELUSION

1. Introduction

In this chapter I explain how the Capgras patient's immediate perception of her loved-one might involve the rich 'impostor' concept. My account takes this to be a basic fact that stands in need of explanation. Accounts based on the Ellis (2007) face processing model, and the delayed-processing account presented in the previous chapter, describe how a putative cognitive impairment produces the conflicting data 'stranger/loved-one' that is thought to underlie the most common form of the Capgras delusion. The issue between explanationists and supporters of an endorsement account is how the conflicting data leads to the impostor belief. According to explanationist accounts, subpersonal processes determine that the loved-one is, in fact, a stranger (or at least odd in some way), and the impostor belief is formed after reflection at the conscious level. According to the endorsement account, subpersonal processes determine that the loved-one is an impostor, and this constitutes the meaning component of the immediate perceptual experience. The belief is formed by accepting the experience as veridical. My account argues for this view, and, therefore, is an endorsement account. What supporters of an endorsement account need to explain is how it is possible for the immediate perceptual experience to acquire such rich content. The following is an attempt to provide such explanation.

2. The phenomenology

If we begin as Gallagher does, with the conscious experience of perception, we need to ascertain the precise nature of the Capgras patient's perceptual experience of perceiving her loved-one prior to post-reflective inferences. What needs to be determined is whether the meaning 'impostor' is part of the experience or something added later on. But obtaining the data is not a simple matter. It is necessary to speculate on the basis of the data we do have about what might be the case and form an hypothesis. A difficulty with this method is that different arguments can be mounted to support different hypotheses based on the same data. For example, in arguing for an explanationist account, Maher

(1988a) describes a woman who believed her head was full of bees. Maher argues that the woman could hear a buzzing sound (of biological origin) in her head and to explain this buzzing sound, hypothesised that it was caused by bees. Her delusional belief acquired its content through an acceptance of this hypothesis. However, it could be argued that the sound the woman heard in her head was not merely a *buzzing* sound but more specifically, a *bee-buzzing* sound. We learn to identify objects and events by the sounds they produce. We hear *thunder*, *a dog barking* or *a baby crying*. At the same time we are immediately aware of the location of the sound. We do not hear a dog-barking sound coming from our garden and consciously hypothesise that the dog-barking sound is in fact being produced by a barking dog, and then hypothesise further, that based on the direction and volume of the sound, there must be a dog in our garden. We directly experience the sound of a dog barking in the garden. If we allow that in this case our immediate conscious perception is 'barking-dog-in garden', then, we might allow that the woman in Maher's case experienced 'buzzing-bees-in-head', and that the delusional belief was formed by accepting the experience as veridical. All we know is that the woman claimed her head was full of bees. We have the belief but not her description of the experience. And knowing the precise nature of the experience is crucial to determining how the belief was formed.

The case of the woman who thought she had bees in her head shows the difference that under-describing the experience can make to the way we interpret the data. Over-describing the experience can also bias our interpretation of data. Langdon and Bayne make the following claim when describing the contents of the Capgras experience.

> While we might allow the perceptual experience could deliver, on first viewing of the loved one, the delusional content, "That man in front of me is a stranger", we can hardly allow that perceptual experience could deliver, on first viewing of the loved one, the delusional content "That man in front of me is an impostor pretending to be my husband". The impostor thought could only have come to mind after some postreflective inference making." (Langdon & Bayne: 2010, 341)

On first glance the rich 'impostor' content, which takes fourteen words to describe, does seem rather too complex to include in the initial experience. But if we analyse it more closely from a phenomenological point of view, their

argument is less convincing. Describing the man as being 'in front of me', which is also cited as part of the 'stranger' experience, is hardly necessary as pretty much everything we see is in front of us. It seems more likely that the information is subliminal, a background assumption applicable to nearly all visual perception. The latter part of the description 'pretending to be my husband' is also problematic. If the patient possesses the concept 'impostor' it is not necessary for her to explain to herself what the concept means. In fact we may assume that it is because the person appears to be pretending to be the husband that the concept 'impostor' has been accessed. By omitting the unnecessary concepts 'in front of me' and 'pretending to be my husband', we can reduce the content of the perceptual experience to 'that man is an impostor'. We might then say that whilst Langdon and Bayne are prepared to allow 'that man is a stranger', they will not allow 'that man is an impostor', claiming that the impostor thought can only come after post-reflective inference making. I suggest that this is not necessarily so. If it is possible to access the concept 'stranger' at a subpersonal level, it might also be possible to access the concept 'impostor' at the sub-personal level.

> There is scarcely daylight between the proposition that this person who looks like my wife (and, indeed, claims to be my wife) is not really my wife and the proposition that this person is an impostor. (Davies: 2009, 74)

Even supposing that the wife remains silent, I suggest that the impostor concept will be accessed if the subpersonal attribution process utilizes the data obtained from context when attributing meaning.

3. Objects, contexts, scenes and scenarios

Whilst it is true that some objects in our environment attract our attention in preference to others, and that from birth humans show a preference for human faces, at the conscious perceptual level the objects we attend to preferentially are perceived as being integrated into a scene or scenario. We do not 'see' our world as a collection of objects or categories that require bringing together in a meaningful way through conscious reflection. What we see are objects imbedded in context. Ganis and Kutas make the following observation.

> Most empirical work on visual object identification has focused on isolated objects. Yet, in our everyday visual environment, objects are embedded in meaningful visual scenes. (Ganis & Kutas: 2003; 123)

Their observation is also true of empirical studies of face recognition. For example, in face recognition studies using Capgras patients, the faces used are usually completely isolated by masking rather than being embedded in meaningful contexts. Further, our ability to describe objects, such as faces, can be misleading because, as Jaspers notes, the act of describing an object isolates it from its context.

> Conscious psychic life is not just an agglomeration of separable and isolated phenomena, but presents a total relational context which is in constant flux and from which we isolate our particular data in the very act of describing them.
> Immediate experience is always *within a total relational context* which we have to dissect if the phenomena are to be described. (Jaspers: 1963, 58)

Using the Ellis diagram, or my own delayed-processing diagram, to describe face processing without reference to the processing of other data, isolates the activity from its context, and it gives the impression that we see 'faces' rather than 'faces-in-context'. In our everyday visual environment, faces are attached to bodies which are embedded in meaningful visual scenarios.

The rapid interpretation of scenes and scenarios is essential if we are to successfully navigate our world. When stressing the importance of evolutionary pressures on the formation of our cognitive processes, Maher (1999) argues that the immediate interpretation of scenes is an important survival mechanism. He gives the example of entering his office and finding there have been minor displacements of items on his desk which he takes to mean simply that his office has been cleaned. He continues with a further example.

> If I find that not only have items been displaced, but that the desk drawer has been broken open and books normally on the bookshelves are now lying in heaps on the floor, I attribute the "meaning" to be that my office has been ransacked. ***This attribution is rapid and forms an essential component of the process of perception.***[50] (Maher: 1999, 558)

"Catching meaning at a glance is a survival instinct" (Greene & Oliva: 2009, 464). The speed and ease with which we process scenes is not disputed. "Real world scenes are incredibly complex and heterogeneous, yet we are able to identify and categorize them effortlessly" (Kravitz *et al.*: 2011, 7322). "With just a glance

[50] My emphasis

at a complex real-world scene, an observer can comprehend a variety of perceptual and semantic information" (Oliva: 2005, 251), and "it is known that humans can understand a real-world scene quickly and accurately, saccading many times per second while scanning a complex scene" (Fei-Fei *et al.*: 2007, 1). We know the process is fast, but not the precise mechanism that enables the interpretation and binding of data. The relationship between object recognition and context is still unclear and "the extent to which object identification is influenced by the background of the scene is still controversial" (Joubert *et al.*: 2007, 3286). Although the extent of the influence is unclear, research demonstrates there is at least some influence. For example, from their investigation of the effects of scene context on rapid object recognition using both behavioural and electrophysiological measures, Sun *et al.* concluded that the results of their studies "are consistent with previous research showing that scene context modulates object processing" (Sun *et al.*: 2011, 40). Research has demonstrated that observers can recognise the content of a complex image in less than 125 ms (Potter: 1975; 1976), and that scene context has some effect on object processing even when glimpsed as briefly as, for example, 80 ms (Davenport & Potter: 2004) and 26 ms (Joubert *et al.*: 2008). But despite the above, it is a huge leap from knowing, for example, that a congruent context increases the speed of object recognition (Palmer: 1975) and that fearful faces are more readily recognised in fearful context (e.g. de Gelder *et al.*: 2006) to the suggestion that context enables the subpersonal processes to make the 'impostor' attribution in the Capgras delusion. Here I wish merely to emphasise the point that research demonstrates that context has at least some influence on object identification at a very early level, that the meaning of a scene can be grasped in milliseconds without conscious reflection, and that it is in the interests of our survival that the binding of object and context at the highest level possible occurs at the rapid, automatic, subpersonal level. How far these ideas can be pushed will be discussed later in this chapter when I will argue that it is not implausible to suggest that the 'impostor' attribution can be made at the automatic, subpersonal level if the object (face) is interpreted in light of its context.

4. The explanationist account

Explanationist accounts of the Capgras delusion claim that the meaning conveyed to consciousness is, at most, that the loved one is a stranger. Taking the commonly used hypothetical example of the Capgras patient who comes downstairs in the morning and sees his wife in the kitchen making his breakfast, it is suggested that the lack of a normal autonomic response causes him to perceive her as a stranger. As Coltheart explains it, the stranger belief enters consciousness fully formed (Coltheart, Menzies & Sutton: 2010). What seems odd in this account is that when attributing meaning, the subpersonal processes clearly fail to take into account extremely salient data. The woman in the kitchen is not merely a stranger. She is a stranger who looks *exactly* like the man's wife. For the husband, it appears that the familiar features are inhabited by an unknown person. Further, she appears to be wearing his wife's clothes, is where his wife would normally be and is doing what his wife normally does. In fact this stranger is clearly pretending to be his wife. The sub-personal mechanisms do have access to at least some of this data (accessing the wife's PIN is part of the story being told), but, according to the explanationist account, this data is not included in meaning attribution. What we are asked to believe is that when face processing produces conflicting data (a) (PIN) identifies the woman as the wife and (b) the arousal orienting response corresponds to recognition of a stranger, the attribution process favours data (b) over data (a). Using Bayesian probabilistic reasoning, Coltheart (Coltheart, Menzies & Sutton: 2010) attempts to explain this preferential treatment (see Chapter 3.). He argues that the 'stranger' hypothesis is more rational than the 'wife' hypothesis. The problem here is that both hypotheses Coltheart offers are clearly inadequate, because whilst each explains one source of data it ignores the other. In selecting the 'stranger' hypothesis the attributional process is not so much explaining the data as merely accepting data (b) in preference to data (a). If 'wife' and 'stranger' are the only possible hypotheses, which is doubtful, the attribution process is left to choose the best of two bad choices, neither of which comes close to explaining the full story. The 'stranger' hypothesis might explain

the lack of normal autonomic response but it gives no explanation for the striking similarity the stranger has to his wife, nor does it give any hint as to why the wife look-alike is in his kitchen cooking his breakfast. This account conflicts with the accounts of Jaspers and Maher, who both stress the holistic nature of the immediate perceptual experience and the importance of the immediate interpretation of scenes rather than isolated objects.

According to the explanationist account, it would appear that salient contextual information is not included in the subpersonal processing that produces our immediate perceptual experience. Rather, having initially attributed meaning to the object of our attention, additional meaning is added, somewhat laboriously at the conscious level. I find this implausible. Conscious cognition is expensive and slow. It seems unlikely that both the slow processing and the overlooking of salient contextual data in the production of our direct experience of our world, would have survived selection processes. A direct grasp of the significance of a scene or scenario, as Maher points out, is part of our survival mechanism. If this is the case, and arguably it is, then just as we can directly perceive our office has been ransacked, the Capgras husband might directly perceive a stranger, who looks exactly like his wife, cooking breakfast in his kitchen. However, even if we allow that to be the case, it is not sufficient to provide the 'impostor' interpretation. To conclude that the wife look-alike/stranger is someone pretending to be the wife (impostor), with the intention to deceive, requires the attributing of motive. It requires an interpretation of the stranger's mental state.

5. Misidentification and intersubjectivity

The Capgras delusion is a delusion of misidentification most commonly occurring with regard to one or more other persons who are thought to be impersonating the loved-one(s). This raises the question of how immediately we can be aware of another person's actions, desires, motivations and the like, and to what extent our understanding is part of direct perception. Theories regarding the mechanisms by which we might understand other people's mental states have generated a plethora of literature. The issues involved are

many and complex, and giving a comprehensive account of the debate is not within the scope of this thesis. The following is an acknowledgement of the two most accepted theories, and a brief account of an alternative theory offered by Gallagher.

The two most accepted theories are theory theory (TT) and simulation theory (ST). According to TT we understand the actions, feelings, desires and so forth of others by accessing and utilizing a theory of human behaviour thought to be either innate or developed during childhood. The theory, often referred to as 'folk psychology', consists of a framework of concepts and a set of rules or laws that connect particular conditions with the behaviour being explained. On this view, interpreting the behaviour of others is essentially an exercise in theoretical reasoning. We create theories about the behaviour of others in much the same way as empirical theories are developed in science. In contrast, ST proposes that we understand the psychology of other persons by using our own psychological processes to simulate theirs. It is based on the idea that when faced with a non-actual situation, such as something that might happen in the future, we imagine ourselves in the situation and observe how we feel and what we would do. We run a 'what if' scenario off-line. In the same way, to understand others we imagine ourselves in their situation and notice how we feel and how we would react. Rather than formulating a theory about the mental states of others, we engage in a form of imaginative role playing to access their psychologies. The theory relies on an underlying assumption that others will feel and react in much the same way as we would in any particular situation.

Gallagher (e.g. Gallagher: 2001; 2009c; 2011; Gallagher & Hutto: 2008) claims both TT and ST are incorrect because they are based on three assumptions which he challenges. The first is the assumption that we cannot directly perceive another's thoughts, feelings or intentions and therefore need some extra cognitive process (theorizing or simulating) that will enable us to infer what they are. Gallagher rejects the idea that other minds are inaccessible and argues that we are able to directly perceive another's intentions, emotions and

dispositions, in their embodied behaviour. The second assumption is that our everyday stance towards others is a third-person, detached, observational stance. Gallagher suggests it is second-person interaction, because mostly when we are interacting with others we are interacting on some project, or in some pre-defined relation. The third assumption is that mentalizing processes constitute our primary and pervasive way of understanding others, rather than being, as Gallagher suggests, the rare and specialised abilities that we develop on the basis of a more embodied approach. As an alternative to TT and ST, Gallagher offers his own theory, 'interaction theory' (IT). Interaction theory has three parts, intersubjective perception, pragmatically contextualized comprehension, and narrative competency. Following Trevarthen (1979), Gallagher calls the first two, primary and secondary intersubjectivity.

i) Primary intersubjectivity

Gallagher's basic claim is that, in most intersubjective situations, we have a direct, pragmatic understanding of another person's intentions because their intentions are explicitly expressed in their embodied actions.[51] He begins, as others do, with the development of the prior knowledge basic to any attribution of intention to others.

> Many of those who argue for the theory or simulation approach acknowledge that for either a theoretical stance or a simulation routine to get off the ground some understanding of the context and behaviour of the other person must be had first; otherwise I would have nothing to simulate or to theorize about. (Gallagher: 2001, 86)

There is empirical evidence showing, that from birth onwards, infants have basic sensory motor capacities that motivate complex interaction between the child and others. For example, the neonate has the ability to pick out a human face from a group of objects, can imitate facial expressions, and can track the eye direction of others. "In addition, infants are capable of discerning emotions and intentions in the postures, movements, facial expressions, gestures, vocal intonation, and actions of others" (Gallagher: 2009c, 293). According to

[51] This reflects Jaspers' claim that we understand an expression directly, without any need for reflection. "We understand in a lightning flash at the very moment of perception" (Jaspers: 1963, 245)

Gallagher, this requires the following prior knowledge that is already implicit in the behaviour of newborns.

> (a) An understanding of what it means to be an experiencing subject; (b) an understanding of what it means that certain kinds of entities (but not others) in the environment are indeed such subjects; and (c) an understanding that in some way these entities are similar to and in other ways different from oneself. (Gallagher: 2001, 86)

Whereas some theorists consider these capabilities to be precursors of theory of mind, Gallagher sees them as clues for an alternative approach. His approach is based on the innate body schema[52], which he describes as an innate system designed for motor control, best understood as a set of pragmatic (action oriented) capabilities embodied in the developing nervous system. "In human infants this system accounts for the possibility of recognizing and imitating other humans" (Gallagher: 2001, 87). Citing Gopnik and Meltzoff (1997, 129), Gallagher says, "we innately map the visually perceived motions of others onto our own kinaesthetic sensations" (Gallagher: 2001, 87). Therefore, as in the case of facial gesture imitation in infants, "one does not require an intermediate theory or simulation to translate between one's own proprioceptive experience of one's face and the visual perception of the other's face" (Gallagher: 2001, 87). This differs from the view of theorists who suggest that infants form a 'plan' or internal representation of what they will do, and then recognise the relationship between their plan to produce the action and the action they perceive in another. Gallagher makes the point that in directly mapping the action of others onto their kinaesthetic sensation, infants are not taking an observational stance. "They are *interacting* with others" (Gallagher: 2009c, 293). Their perception-based understanding is not a form of mind-reading.

> Thus, before we are in a position to theorize, simulate, explain or predict mental states in others, we are already in a position to interact with and to understand others in terms of their expressions, gestures, intentions, and emotions, and how they act towards ourselves and others. (Gallagher & Hutto: 2008, 22)

Importantly, Gallagher suggests that primary intersubjectivity is not something we leave behind as we mature. It is something we rely on across all face-to-face intersubjective experience, even after we have developed more sophisticated abilities.

[52] For Gallagher's description of the body schema see Chapter 1.

ii) Secondary Intersubjectivity

At about one year, the child not only sees movement, gestures, actions and the like, as being embodied, but now also comes to see them as being embedded in the world. "In secondary intersubjectivity the world begins to do some of the work as we try to understand others" (Gallagher: 2009c, 294). The world around us and the things in it set the stage for carrying out certain actions, and the pragmatic and social situations in which the child encounter others helps her make sense of their behaviour. It gives access to the other's intentions in the immediate environment, here and now. For example, by eighteen months children are able to comprehend and re-enact to completion, goal-directed behaviour that another has failed to complete.

> By the end of the first year of life, infants have a non-mentalizing, perceptually based, embodied and pragmatic understanding of the intentions and dispositions of other persons" (Gallagher: 2011, 62).

In infancy our understanding of the behaviour, intentions and feeling of others is second-person, non-inferential, embodied and embedded, and according to Gallagher, this stance remains our primary mode in adulthood. He claims that we never see the environment, the situation or the pragmatic context as being without meaning either in regard to our own possibilities for action or those of others.

> Rather than making an inference to what the other person is intending by starting with bodily movements, and moving thence to the level of mental events, we see actions as meaningful in the context of the physical and intersubjective environment" (Gallagher & Hutto: 2008, 24)

Our perception of another is never of an entity existing outside a situation, but of one in a pragmatic context which helps us understand their intentions (Gallagher: 2008b).

iii) Narrative competency

After the age of two, language enables the child to understand more complex actions and intentions that extend over time, and this helps the development of narrative as a framework for understanding the actions and intentions of others.

> Starting in a preliminary way around two years, and fostered by the stories that we read to children, narrative builds and expands on secondary intersubjectivity and starts to provide more subtle and sophisticated ways of framing the meaning of the other's intentions and actions. (Gallagher: 2009c, 294)

The development of narrative competency is developed when children are exposed to, and engage in story-telling practices with their care-givers. Stories about those who act for reasons (folk psychological narratives) provide exemplars on which children build understanding. During story telling (or reading) children are encouraged to explain why characters behave in certain ways, their motivations, feeling, wishes or desires, with the stories increasing in complexity as the children mature. Gallagher suggests it is in this way that children develop an implicit, practical understanding of *how* to make sense of persons as those who act for reasons. He notes that the proposal is consistent with a number of recent empirical studies that have established that there are important links between narrative ability and our capacity to understand others (Gallagher & Hutto: 2008, 29). Different kinds of narratives familiarize children with sets of characters and with a wide range of ordinary and extraordinary situations, from which they learn that actions need to be appropriate to characters and situations. Through narrative children become accustomed to the standard scripts, characters and scenarios that typify their social interaction and their cultural heritage, and they learn the norms associated with the social roles that pervade their everyday environment. Important to developing narrative competency is the child's ability to imaginatively identify with, and respond emotionally to the narrative, at least to some degree, in the same way they do in basic social engagement. Children learn to *be-in-the-world* of the story.

Gallagher (2009c) describes two hypotheses in regard to narrative competency, the *implicit framing hypothesis*, and the *narrative practice hypothesis.* The first hypothesis states that through developing narrative competency we come to implicitly make sense of our own and others' actions in narrative frameworks. "Our perception and understanding of the behaviour of others comes to be pre-reflectively shaped by narrative" (Gallagher: 2009c, 294). This means that we implicitly understand that the behaviour of others is part of a story-line, and we

interpret their behaviour accordingly. The second hypothesis states that narrative provides the concepts that are basic to folk psychological practice. If a situation is puzzling, or we have some other reason to take a third-person perspective, our narrative can become reflectively folk psychological narrative. That is, we can use folk-psychological narrative to take a mind-reading stance.

> Folk psychology is needed only in rare cases where we are not already familiar with the other person's story, or are perplexed by another's actions. For "When things 'are as they should be', the narratives of folk psychology are unnecessary" (Bruner: 1990, 40). (Gallagher & Hutto: 2008, 30)

One might argue that to acquire narrative competency requires having a theory of mind. Gallagher refutes this suggestion, arguing it is possible to understand the major events in a drama without being able to understand fully all the reasons for another's actions. That is, we can have some sense of what is going on without having access to the other's inner life characterised as a series of causally efficacious mental states. We might say we can grasp the gist of what is going on in a glance.

> Coming to understand another's reasons should not be understood as designating their discrete 'mental states' but their attitudes and responses as whole situated persons. I encounter the other person, not abstracted from their circumstances, but in the middle of something that has a beginning and that is going somewhere. I see them in the framework of a story in which either I have a part to play or I don't. The narrative is not primarily about what is 'going on in their heads'; it's about the events going on in the world around them, the world that we share with them, the events in their *lives* and the way they understand and respond to such events. (Gallagher & Hutto: 2008, 33-34)

For Gallagher, mentalizing as characterized by explanation and prediction, such as posited by TT and ST, is not our usual stance. Explanation and prediction "are specialized and relatively rare modes of understanding others" (Gallagher: 2001, 94). Our pervasive stance is second-person pragmatic engagement with others in which we implicitly understand their behaviour from their embodied actions and the rich worldly contexts within which they act, "contexts that operates as scaffolds for the meaning and significance of actions and expressive movements" (Gallagher & Hutto: 2008, 34).

6. The Attribution Process

Both TT and ST share the assumption that mental states are unobservable and therefore mental state attribution requires an extra psychological step that goes

beyond perception. What this means depends on how one defines 'perception'. On Raftopolous' definition, all cognitive steps are beyond perception. On Fodor's definition, it is all cognition beyond basic categories. The 'common sense' view generally accepts that, in most situations, we can grasp the intention of others from their behaviour. And "Certain phenomenological approaches depend heavily on the idea that we have a direct grasp of the other person's intentions, feelings, etc." (Gallagher: 2008b, 535).

> When I see the other's action or gesture, I see (I *immediately perceive*) the meaning in the action or gesture. I see the joy, I see the anger, or I see the intentions in the face or in the posture or in the gesture or action of others. (Gallagher; 2008b, 542)

The issue here is not where we draw the interface between perception and other cognitive processes but rather where we draw the interface between the subpersonal processes and conscious awareness. The question is how much of the attribution process is performed at the subpersonal level and how much is performed at the conscious level. Neither TT nor ST theorists claim that the processes they describe occur at the conscious level. To be compatible with the view that the content of the immediate perceptual experience contains an interpretation of the mental states of others, it is only necessary for the processes to run *fast* enough to enable the output (e.g. 'impostor') to provide the content of the immediate perceptual experience. Therefore, my account is neutral with regards the mechanism by which we interpret the thoughts, feelings and intention of others. However, I admit to a leaning towards Gallagher's theory, primarily on evolutionary grounds. If he is correct in saying that other minds can be directly accessed by mapping their embodied actions onto our own kinaesthetic sensations, this mechanism would better suit survival demands than ST and TT. Both ST and TT require an extra cognitive step, making them slower and more expensive in terms of cognitive effort. As both speed and conservation of cognitive effort enhance survival, Gallagher's mechanism is the one most likely to have evolved because of its greater survival benefits.

7. The 'impostor' attribution

Box-and-arrow face processing diagrams, such as the delayed-processing and the Ellis diagrams, show how additional information might be added and integrated in cognitive steps. But when it comes to explaining the contents of the immediate perceptual experience, these diagrams only play a partial role because they only describe the processing of visual data provided by a face. [53] When recognising a person, information from the other senses is integrated (Young, Ellis & Szulecka: 1990), and the fact that we can fail to recognise a familiar person in an unfamiliar context demonstrates that context also plays a role in person recognition.

Putting aside the issues of where and how the various components of our visual experience are bound together, and allowing that both research and argument from evolution support the idea that people are able to immediately grasp the gist of a scene, and accepting, as does Maher, that we are able to grasp the gist of the meaning of a scene, we can ask what we might sensibly allow to be included in the contents of the immediate experience of the Capgras patient. Again using the hypothetical example of the wife preparing breakfast, it can be argued that the Capgras patient should be able to rapidly identify the context (kitchen), his wife's face, body, clothing, posture and any other visual features associated with his wife. Drawing on experience, identifying his wife's behaviour (preparing breakfast) should also be a rapid process. Having grasped the gist of the scene, the question is whether or not he is also able to grasp the gist of the meaning of the scene. The scenario that needs interpreting is one in which a stranger who looks exactly like his wife and who is dressed like her, is in his kitchen doing what his wife normally does. That is, there is a stranger in his kitchen who is impersonating his wife. My contention is that the gist of the meaning would be

[53] I am using the causal chain account of cognitive processing here and for my diagrams in other chapters, as it is the one used by Ellis. I acknowledge that cognition may not operate on a computer model and it is my personal view that it does not. However I do not think a distribution network account, for example, would substantially alter my argument. My argument is that rapid, automatic, subpersonal mechanisms process the outputs of various data processing systems to produce complex interpretations of our world. The precise nature of the mechanism involved is not the issue.

immediately apparent, that is, that the woman's intention to dupe him into believing she is his wife can be grasped from the way in which she is dressed, her general behaviour, and her behaviour towards him, in the given context.

The fact that the patient does not reject the interpretation of the scene as implausible indicates there is a second factor. As briefly described in the introduction to this section, the second factor in my account is the decoupling of a conflict monitoring system from executive control, which is the normal mechanism we employ to enter an alternative reality. Before presenting a detailed account of the second factor, I will argue that the 'impostor' concept that provides the content of the gist of the meaning of the scene is primarily developed through acquaintance with fiction. The following two chapters are devoted to arguing the case, beginning with a discussion of the relationship between delusions and alternative realities.

CHAPTER 8: ALTERNATIVE REALITIES AND DELUSIONS

1. Introduction

In this chapter I argue that the 'impostor' concept that provides the content of the Capgras delusion is drawn from the patient's conception of an alternative reality. Cognitive scientists generally assume that delusional patients share the same reality as the non-delusional, but for the delusional patient something goes awry in the process of forming beliefs about that reality. Their emphasis is on describing the belief-formation process, the point at which it fails, and the way in which it fails. Phenomenologists, such as Jaspers, Sass and Gallagher, contend that delusional patients do not share the same common reality shared by the non-delusional, but experience an alternative reality on which their beliefs are based. Their emphasis is on describing the nature of the alternative reality that the patients experience. In keeping with the phenomenological view, the rich account of the contents of the immediate perceptual experience for which I am arguing, proposes that the concepts used to attribute meaning to sense data are drawn from different, circumscribed realities or worlds, each with its own web of beliefs and set of logical possibilities. The account is drawn from Shaun Gallagher's description of multiple realities (see Chapter 1.) and his account of intersubjectivity (see Chapter 7.). It also draws on Kendall Walton's description of make-believe and the construction of imaginary worlds, an account of which is given in this chapter. I begin with a brief overview of the phenomenological view of alternative realities and conclude with a discussion of the fictional and non-fictional nature of the 'impostor' concept.

2. Delusions and alternative realities

Jaspers introduced the idea that deluded patients experience a different reality to that of everyday life when he described the delusional experience of reality as one "in which the environment offers a *world of new meanings*" (Jaspers: 1963, 99). He describes the delusional atmosphere, which he considers one of the main characteristics of primary delusions, in the following way.

> Patients feel uncanny and that there is something suspicious afoot. Everything gets a *new meaning.* The environment is somehow different – not to a gross degree - perception is unaltered in itself but there is some change which envelops everything with a subtle, pervasive and strangely uncertain light. (Jaspers: 1963, 98)

For Jaspers, the world of new meanings the patient experiences render it incomprehensible to those experiencing everyday or 'normal' reality. Sass is less pessimistic on this point but supports the idea of an alternative reality.

> Many schizophrenic patients seem to experience their delusions and hallucinations as having a special quality or feel that sets these apart from their "real" beliefs and perceptions, or from reality as experienced by the "normal" person. (Sass: 1994, 3)

As further indication of the alternate nature of the delusional world, he points to the phenomena of double-bookkeeping, a patient's tendency to firmly assert a delusional belief whilst at the same time behaving in keeping with everyday reality.

> It is remarkable to what extent even the most disturbed schizophrenics may retain, even at the height of their psychotic periods, a quite accurate sense of what would generally be considered to be their objective or actual circumstances. (Sass: 1994, 21)

Sass considers it a mistake to view the patient's assertions as (failed) attempts to refer to the same kind of common sense world people normally experience. Although both Jaspers and Sass are describing the delusional world of the schizophrenic, there is reason to believe Capgras patients also experience alternative realities. According to Christodoulou (1977; 1986), Capgras patients commonly report experiencing pervasive feelings of strangeness, and feelings that everything is somehow unreal or unfamiliar. And Stone and Young note that "this has been a persistent finding ever since the original paper by Capgras and Reboul-Lachaux (1923)" (Stone & Young; 1997, 337).

To enter an alternative reality is, in itself, not to be delusional. As Gallagher points out, the experiencing subject does not live in one unified world of meaning, but in multiple realities, sub-universes or finite provinces of meaning. These are the worlds of dreams, cinema, theatre, literature and virtual reality. Many worlds, such as theatre, are consensual worlds in which the rules determining what is logically possible within each world are agreed upon. They may be shared by anyone who wishes to participate by accepting the agreed rules. Other worlds, such as those of dreams and daydreams, are private and

idiosyncratic. According to Gallagher, when entering an alternative reality, we adopt a set of beliefs and values that are not commensurate with everyday reality. But more significantly, in doing so, we undergo an existential change that allows us to be *in-the-world* of the particular reality. If delusions are experiences of alternative realities, it might explain the apparent bizarre content of many delusions. And he notes the common themes in particular types of delusions may be drawn from themes in our shared culture.

> The MR hypothesis suggests that we examine certain shared features of our cultural environments (our other multiple realities) that may lead to the particular types found in delusion. Why there are typical kinds of delusions may call for the same kind of answer that we would give to the question of why there are typical scenarios developed in pretend games and imaginary play in childhood, or 'universal' literary themes found in novels, plays and other media. (Gallagher: 2009a, 262)

He proposes further, that possible areas of brain damage might open the door to a delusional reality of a particular kind. In the previous chapter I described how the meaning 'impostor' might be accessed as a way of resolving the conflict between the wife/stranger attributions that results from an impairment of face encoding. In this chapter I suggest that, for most people, the concept of one person pretending to be another is acquired during engagement in childhood games of make-believe and stories, and later, expanded through the experience of cinema, theatre and literature. In conjunction with the second factor, which will be explained in detail in Chapter 10, I suggest further, that the original impairment ultimately causes the patient to enter the alternative reality of a fictional world. My account is in keeping with Gallagher's suggestion that some kind of brain damage might open a door to an alternate reality that has its roots in *being-in-the-world* of pretend games and imaginary play in childhood, and in the patient's cultural environment. In the following section I will focus on the general nature of alternative worlds and their construction before discussing how the content of the Capgras delusion might be acquired in accordance with this view.

3. Creating imaginary worlds

If we accept the proposition that novels, films, theatre and games are alternative worlds that take us away from everyday reality, questions arise about the

manner in which these imaginative worlds are formed. One approach to answering the questions is to classify the worlds as make-believe, or pretend worlds, and then consider how these imaginary worlds are created and how we engage in them. The literature on make-believe and imagination is vast. As there is insufficient room here to consider all the available accounts, I will restrict myself to one—that offered by Kendall Walton in his seminal work '*Mimesis and Make-Believe'* (1990). As it is his broad suggestion that imagination involves alternative realities, not the fine detail of his exposition that is important to my account, the work of several other writers who hold the same general view could have been used.

Walton's interest is in the representational arts of which he considers novels, stories, figurative painting and sculpture, theatre and film to be paradigmatic instances (Walton: 1991). These cover the alternative realities described by Gallagher and there seems no reason why we could not include being *in-the-world* of a painting or sculpture in his theory of multiple realities. As the representational arts include painting and sculpture, Walton resists the various linguistic approaches to explaining make-believe. His emphasis is on the *participatory* nature of imagination in the appreciation of representational art rather than on the use of imagination in its creation. He claims we do not merely stand back and observe fictional worlds from the outside but enter the world of the work in an imaginative act. In explaining his theory he draws attention to children's active participation in make-believe and the way in which they belong to the world of their games, and he suggests that the way adults relate to representational art is analogous with children's games.

> I claim, however that appreciators use paintings and novels as props in games of make-believe, much as children use dolls and toy trucks, and that appreciators participate in these games. (Walton: 1991, 380)

As well as explaining how we appreciate representational art, Walton claims to articulate a theory of make-believe likely to have many other applications as well. In the following chapter, I propose to show that using it to explain certain puzzling characteristics of the Capgras delusion is one of the other applications.

Make-believe, explained in terms of imagination, constitutes the core of Walton's theory and he uses it to explain how it enables an appreciator to enter the world of a particular work. We enter the work by *imagining it is real*. He notes that when a reader abandons herself in a novel it seems she is convinced momentarily and partially at least about the characters' existence. "Otherwise why would she be moved by their predicament?" he asks. "Why would one even be interested enough to bother reading the novel?" (Walton: 1990, 7). Walton's description of the sense that the characters in a novel exist in some real way is evident in the following exchange between Leigh Sales and the author Jonathan Franzen.

> **Sales:** As a reader, if the characters are particularly well drawn, I sometimes feel like they're still living their lives somewhere out there and I just don't have a window on that anymore after I have finished the book.
> **Franzen:** As a reader I absolutely feel that way. I really do believe, just like you said, they are out there in the world. They have some reality. (ABC Lateline: 2010)

According to Walton, the experience of being "caught up in a story" and emotionally involved in it is central to the appreciation of much fiction, and "It is extraordinarily tempting to suppose that when one is caught up in a story, one loses touch with reality, temporarily, and actually believes in the fiction" (Walton: 1990, 6). But, of course, the normal appreciator does not *really* believe in the fiction. Therefore, Walton suggests that the central metaphysical problem concerning fiction is mirrored in the very experience of appreciation, because to appreciate fiction, which we know is not true, requires that in some kind of way we accept that it is.

Walton's theory employs the concepts of *fictional truths*, *props, principles of generation* and *fictional worlds.* A *fictional truth* in general is whatever is the case in a fictional world, that is, in the world of a game of make-believe, a dream or daydream, or a representational work of art. It is a fictional truth that Tinkerbell is a fairy and that Ophelia drowned in the river.

> Propositions that are "true in a fictional world", or fictional, are propositions that, in a given social context, are to be imagined as true. (Walton: 1991, 380)

Props are generators of fictional truths because they stimulate the imagination. A doll is a prop because it prompts the child to imagine she is holding a baby.

Representations are also props because their function is to stimulate imagination in a game of make-believe. For example, words can readily be employed to prescribe imaginings. When a collection of verbal or written words prescribe imaginings, it is a prop. The specific fictional truths generated by a particular set of props operative in a relevant social context is governed by the *principle of generation.* Walton gives as an example a group of children who, while playing in the woods, pretend that tree stumps are bears. The principle of generation in the children's game of make-believe is the basic stipulation 'let all stumps be bears'. To participate in the game rather than merely be an onlooker, each child has to agree that stumps are bears. Participants consider themselves constrained to imagine propositions that are fictional in the game of make-believe, that is, they consider that the principles of generation apply to *them*. "A principle is in force in a particular context if it is understood in that context that, given such-and-such circumstances, so and so is to be imagined" (Walton: 1990, 40-42). Not all principles of generation relate directly to objects in the world, such as tree stumps and dolls. Some imaginings, like private daydreams, are generated internally by a rule that says, for example, 'imagine winning the lottery'. These rules are occurrent and idiosyncratic. In contrast, participation in social games is by agreement and often repetitive. In repetitive games of make-believe, the rules do not have to be set up each time a particular game is played as they can be learned and come to operate as a type of game language.

> Insofar as it is the object's recognized function to be a prop in certain kinds of games, the principles are likely to seem natural, to be accepted automatically, to be internalized, and the prescribed imaginings are likely to occur spontaneously. (Walton: 1990, 53)

We might think, for example, that the generating principle for reading a novel might be something like 'pretend everything I read is about real people in real situations'. The text of the novel would then serve as a prop that prescribes the particular imaginings in this game of make-believe. The *generating principle* would apply to every novel with the *prop* dictating the particular contents of the imaginings of each particular novel. Unlike the principle 'let all stumps be bears', which is proposed and then agreed to by each participating player in the game, the generating principles of novels, films and plays are internalized. We do not

consciously agree to accept a proposition to 'pretend' before we participate in the game.

Fictional worlds are simply created by clusters of fictional truths that relate to each other. Each fictional world is associated with a particular class or cluster of propositions that are fictional to that world. They differ from possible worlds in that possible worlds (as normally construed) are necessarily both possible and complete, whereas fictional worlds are sometimes impossible and usually incomplete (Walton: 1990). According to Walton, truth in a fictional world is not the same as truth in the real world. In fact it is not a truth, and we should resist the temptation to regard both as a species of truth. Some philosophers suggest that it is discourse itself that creates our 'reality' and there is little difference between this and a novel. The 'real world' is merely a fancy name for just another fictional one. Walton disagrees.

> Reality is reality and facts are facts, however they are to be understood, and what is the case obviously does differ from what is not the case, even if the difference is somewhat conventional, culturally specific, dependent on this or relative to that, or whatever. Walton: 1990, 110)

He points to the fact that every thought or piece of discourse that aspires to truth has a reality independent of itself to which it has to answer, whatever role sentient beings might have had in constructing that reality. A fictional world corresponding to a work of fiction is not independent of the work. The fictional truths in a fictional world are only true in relation to the generating principle and the props. Fictional truths are only true within a game we have agreed to play. Walton offers the following brief summary of his theory.

> A prop is something which, by virtue of conditional principles of generation, mandates imaginings. Propositions whose imaginings are mandated are fictional, and the fact that a given proposition is fictional is a fictional truth. Fictional worlds are associated with collections of fictional truths; what is fictional in a given world – the world of a game of make-believe, for example, or that of a representational work of art. This, in brief outline, is the skeleton of my theory. (Walton: 1990, 69)

To this brief summary we might add that the kind of existential shift described by Gallagher is necessary to enable a participator to enter an imaginary world and experience it 'as if' it were real. And the more absorbed the participator becomes in her imaginary world, the greater the shift, and the more 'real' it becomes to her.

4. Make-believe and the Capgras delusion

Before discussing the Capgras delusion and the proposal that the 'impostor' attribution is sourced from an alternative reality, it needs to be acknowledged that impostors do exist in the real world. There are abundant real life examples of one person posing as another with the intention to deceive. These include thieves, who impersonate technicians or council workers to gain access to people's homes, those posing as owners of stolen credit cards, scam operators and the like, decoys used by celebrities to misdirect the paparazzi and, decoys used by presidents and dictators for security reasons. The tendency to think of impostors as only inhabiting fictional worlds is demonstrated by Reimer, who in discussing the Capgras delusion says "In order to embrace the hypothesis, the patient must, after all, give up his previously held view that impostors do not exist" (Reimer:2009, 675). She likens impostors to ghosts, goblins, witches, warlocks, poltergeists, doppelgangers, UFO's, Big Foot and the Loch Ness Monster, all of which she claims, are deserving of scepticism.

> Although scientist and philosophers (and educated persons more generally) typically regard all such phenomena as equally incredible, that does not mean that they are so regarded by persons *generally.* Many people believe in the possibility – if not the actuality – of such spectacular things. (Reimer: 2009, 676)

The point Reimer stresses is that we should not be too quick in supposing that the Capgras patient originally believes impostors do not exist, and therefore, has to overcome her scepticism to accept the hypothesis. Reimer argues that if the patient already believes impostors are real, the impostor hypothesis would require less disruption to the patient's existing set of beliefs than would the 'brain-damage' hypothesis.

Reimer's mistaken claim that impostor do not exist may stem from the fact that very few people develop the 'impostor' concept from exposure to real impostors. Beginning in childhood, the concept is usually developed through engaging in fictional narratives, such as *Little Red Riding Hood*, and consolidated by further engagement in fiction. The expansion of the concept to include non-fictional impostors is also unlikely to be the result of personal experience.

Impostors exist in both fictional worlds and the real world, and as many fictional worlds are drawn from real world situations, it is difficult in the case of the Capgras delusion, to determine the nature of the world from which the concept is sourced by merely looking at the content of the delusion. However, it is clear that sometimes a delusion has its roots in a fictional world. As Stone and Young note, "Attempts to account for the substitution may involve the secret service, Martians, robots, clones, or a frank admission of its inexplicability (Stone & Young: 1997, 333). In the case of Martians, robots and clones, Reimer appears to be right in suggesting that the impostor the patient perceives is a fictional one. But it is not clear that all explanations are drawn from fiction. A patient's claim that her mother and the impostor killed her father to enable them to start a new life together (Brighetti: 2007), could have its roots in the plot of *Hamlet* or from a news report of a real event. Claiming his parents and siblings were shot by Chinese communists (Doran: 1990) may well be drawn from a real life event, but it is unlikely the same is true for a patient's claim that her family were taken out front of the hospital and shot by the hospital staff (Tally & Michels: 2009). Cases where the spouse or parents are reported to have been killed in a car crash (Ellis & Lewis: 2001) provide plausible real world explanations for the loved-ones' absence, but the scenario is also common in fictional worlds. It is important to note that the *amount* of content varies between patients. As Stone and Young (1997) observe, some patients can give no explanation for the presence of the impostor. Others create quite elaborate stories to explain the impostor's presence, and/or the loved-one's disappearance. I suggest there are two different processes involved here. The first is that the patient's immediate perceptual experience, which provides the content 'a stranger is pretending to be my loved-one, with the intention to deceive'. As the content is experiential, it is not accompanied by either reasoning or explanation; hence the inexplicable nature of the experience described by some patients. The first process is in keeping with phenomenologist's 'experiential' explanation of delusions. The second process is that of elaboration. Not content to accept the inexplicable, some patients attempt to explain the presence of the impostor with a cluster of fictional truths,

thus creating a fictional world around the anomalous experience. The second process is in keeping with Maher's 'explanationist' account of delusions. Because there is, in fact, no impostor, all elaborations are, of necessity, fictional. The point I want to make here is that some elaborations, such as those involving Martians and robots, are clearly based on fictional narratives, whilst others could be based on real life events. Consequently, we cannot use the content of the delusion to demonstrate that the meaning component of the immediate perceptual experience is drawn from a fictional world. However, I suggest there are other reasons to think that they are, and these will be discussed in the next chapter.

It might be argued that children have multiple experiences of impostors in the real world when, for example, their father plays Santa Claus, or friends pretend to be fairies or monsters. I suggest that these games of pretence do not fully capture the impostor concept. The difference between a person who is pretending to be someone else (a pretender) and an impostor is one of intention. The intention of the impostor is to genuinely deceive. We do not call the actor playing the role of King Lear an impostor because he is not attempting to trick us into sincerely believing that he *is* King Lear. He is inviting us to join a game of make-believe. Likewise, the child who is pretending to be a fairy or a monster is inviting others to join her game of make-believe. The second difference is the objective or purpose of the impersonation. Impostors impersonate for personal gain or profit, or to protect themselves or the innocent from villainy. The situation often involves the potential death of a victim or a villain, both in fiction and the real world. It is essential to the realization of the impostor's motives that she not be unmasked. The pretender's objective is to create fun or entertainment through game playing. The father who is pretending to play Santa Claus is doing so for the child's enjoyment, not for personal gain or to protect the child from harm. His objective will be realized whether or not he is unmasked. Children understand that the wolf is not inviting Little Red Riding Hood to join him in a game of make-believe. Rather, they understand he is attempting to trick her into sincerely believing that he is her grandmother for

his own dreadful purposes. There is no fun planned for Little Red Riding Hood. Neither does the bodyguard who detects an impostor attempting to breach security view it as an invitation to engage in a game of make-believe. Unlike the motives of the actor playing King Lear, she understands the impostor intends to seriously deceive for villainous purposes. Arguably, very few children are regularly exposed to the type of deception perpetrated by real impostors.

The above discussion concerns the formation of the 'impostor' concept, and I have argued that, for most people, the concept is formed through exposure to fiction, and, therefore, that the 'impostor' held in memory is primarily a fictional character. However, the Capgras delusion does not involve the *forming* of the concept, but rather, the *applying* of the concept. More specifically, it involves applying the concept of a (fictional) impostor to a real world person. As will be described in Chapter 10, this causes a cross-world conflict, which stimulates the activation of the mechanism that enables people to enter an alternative reality. As a result, the patient is able to engage in the impostor fiction. There is no exact everyday analogy for the Capgras delusion because the sequence of processes that result in the delusion are instigated by a neuropsychological impairment, but it can be likened to our experience of theatre. When at the theatre, we may initially see the actors as themselves, particularly if they are well known to us. As we become more absorbed in the story, we begin to 'see' them as the characters they are portraying. We no longer see Sir Laurence Olivier. We 'see' Hamlet. However, to do so, we need to enter an alternative reality, which, according to Gallagher, requires an existential shift. Similarly, when the Capgras patient applies the impostor concept to her loved-one, she 'sees' a real person as a fictional character. As with the theatre goer, it requires an existential shift from the real to a fictional world. The difference is that when we are at the theatre, we know we are engaged in fiction, but the Capgras patient does not. The mechanism that I suggest enables us to make the necessary shift to an alternative reality, and how this applies to the Capgras delusion, will be described in Chapter 10.

As noted previously, the contents of some Capgras delusions, such as robots and Martians, clearly are sourced from fiction. The source of the contents of other Capgras delusions is less clear. It is my contention that all Capgras impostors are fictional, and that there are reasons for thinking so, other than the bizarre content of some delusions. These reasons relate to the puzzling characteristics of the delusion. If the patient 'sees' her loved-one from the perspective of fiction, it could explain these puzzling characteristics. I will argue this point in the next chapter.

CHAPTER 9: EXPLAINING UNUSUAL CHARACTERISTICS

1. Introduction

In this chapter I use an alternative-realities hypothesis to explain some of the puzzling characteristics of the Capgras delusions. One puzzling characteristic is that most patients respond to the contents of their delusional belief with a passivity that seems inappropriate to a sincerely held belief, whilst others respond actively, even violently to the content. Another is the circumscribed nature of the delusion, which enables patients to hold a sub-set of beliefs that do not cohere with their general web of beliefs. Circumscription is expressed in the phenomenon of double-bookkeeping, which is the patient's tendency to firmly hold a delusional belief whilst acting in keeping with everyday reality. Patients also exhibit a puzzling resistance to revising their delusional belief in the face of counter-argument and evidence to the contrary. In addition, the content of the elaborations that sometimes accompany the delusion is often bizarre.

As part of the explanation of the unusual characteristics, I shall describe the similarities and differences between the non-deluded who sometimes confuse reality and fiction when playing out a fiction in real world, and the Capgras patients who confuse a fictional character with a real person. I will conclude that whilst the alternative realities hypothesis gives a good account of the unusual characteristics listed above, to fully explain the presence of the Capgras delusion, a second factor is needed.

2. Passivity and aggression

Despite the strength of their delusional belief, patients frequently do not react in the way one would expect them to react if they sincerely held the belief. For example, Capgras patients frequently exhibit a puzzling lack of concern regarding the whereabouts of their loved one(s). The lack of curiosity about the fate of the replaced person is a common feature of the Capgras delusion and

therefore, it needs to be explained. One way of explaining the passivity is to reject the doxastic status of delusions.

> One has to ask whether the patients really believe their claims, or whether they are meant as some form of metaphor, or even constitute what Berrios (1991) calls 'empty speech acts that disguise themselves as beliefs'. This raises difficult issues concerning the nature of beliefs, and the criteria for judging sincerity and conviction. One simple criterion would be to ask what patients *do* about their expressed delusional beliefs. (Young: 2000, 52-53)

On this simple criterion, patients who do not act on their stated belief would be regarded as not holding that belief. Several attempts have been made to maintain the doxastic status of delusions by offering alternative explanations for the passivity exhibited by patients. Bayne and Pacherie (2005) note that action is not caused by cognitive states alone, but by cognitive states in conjunction with motivational states. Thus patients may lack the motivation to act, or be motivated *not* to act.

> There is no single action that is *intrinsically* appropriate to a belief. An action is only appropriate or inappropriate to a belief relative to some project that the agent has. (Stone & Young: 1997, 353)

Patients who are aware of the absurdity of their beliefs may not be motivated to pursue the matter because they fear they will not be believed. When asked why he did not report his wife's absence to the police, one patient explained that he did not think the police would believe him because his wife had been replaced by an alien (see Bortolotti: 2005). Stone and Young (1997) claim even when the action is as simple as asserting their belief, it will cause distress and recriminations, and further, that patients who do not report missing persons to the police may choose not to do so to avoid drawing unwanted attention from the medical profession. "It is surely a perfectly understandable project for the deluded person to try to avoid these consequences" (Stone & Young: 1997, 353).

Even if we accept that the failure to take action is understandable, and I am not convinced that it is, what still need explaining are the many cases in which patients not only fail to seek their missing loved-ones, but also fail to grieve for them. More problematic are cases in which patients form a strong positive relationship with the impostor. One review of cases (Wallis: 1986) shows around 30% of patients were friendly towards the impostor, and in one case, a man expressed gratitude for having a substitute provided (Alexander, Stuss &

Benson: 1979). This is not how we would expect normal persons to behave towards someone who was impersonating their missing spouse.

If we take the degree of action as a measure of the sincerity and conviction of a belief, the passivity Capgras patients often exhibit in relation to their delusion might bring the doxastic status of delusions into question. But the issue is more complex than that because not all Capgras patients are passive. In a well known case, one patient accused his stepfather of being a robot and decapitated him in an attempt to find the batteries (Blount: 1986). In another, a woman killed her mother after five years of suffering from the delusion her parents were impostors (Silva et al: 1994).

> It seems likely that the nature of the CS delusions and associated hostility heighten the potential for violence directed towards the misidentified person. This indicates an actual danger of violence from individuals with CS – a danger underscored by the relative frequency of reported severely violent acts, including homicide. (Bourget & Whitehurst: 2004, 723)

Any satisfactory explanation of the Capgras delusion needs to account for both the high number of cases in which patients do not act on their beliefs, and those cases in which patients do act, sometimes with extreme violence.

The range of responses, from passive to violent, exhibited by Capgras patients, might be explained by the hypothesis that the impostor attribution is drawn from a fictional world entertained by the patient. That is, the Capgras patient's impostor is a fictional character that the patient takes to be real. In the following chapter I will argue that, when viewing her loved-one, the second factor causes the patient to enter an alternative reality involuntarily. According to Gallagher, when we enter an alternative reality we adopt a set of beliefs and values that are not commensurate with everyday reality. As the rules, logical possibilities, values and beliefs in imaginary worlds differ from those in the real world, the way in which we respond to characters and events in fictional worlds differs from the way in which we respond to them in the real world. One difference is the level of activity/passivity with which we respond. When we enter the world of a film, play or novel, we accept that we cannot influence the narrative in the way we can in the real world. As a woman seated behind me at the opera, la

Traviata, lamented when it ended, "Oh! She always *dies* at the end of la Traviata." It is true that she always does, and as members of the audience there is nothing we can do about it. We might be *in-the-world* of the fiction, but we are in it in the role of a passive observer. Arguably, repeated experience of *being-in-the-world* of films, plays and novels, creates a strong link between passivity and fictional characters and events. If the Capgras patient's 'impostor' is fictional, as argued in the previous chapter, it is probable that the impostor she perceives is strongly associated with non-action, and that this constrains her response to both the impostor and to the narrative created in response to the impostor's presence.

The above proposition might help explain the passivity Capgras patients often exhibit, but it does not explain the patient's active, and sometimes aggressive, response to the presence of the impostor. There are three components to the alternative-reality explanation of more active responses. The first relates to the patient's level of absorption in the fiction. The degree to which we become absorbed in a fictional world varies, and as our absorption increases so does our emotional engagement. We also become increasingly removed from our awareness of the real world. It is possible that Capgras patients, who respond actively, even aggressively, are those who have become more absorbed in the fiction than those who respond passively. They have become totally caught up in the story. The second component concerns the reality of our real circumstances. When totally absorbed in a fictional world, we may lose awareness of the real world, but this does not cause it to disappear. Our real circumstances are ever present, and normally, any interruption to our focus on the fictional narrative will trigger our re-entry into the real world. The difficulty for the Capgras patient is that rather than the fiction being set *against* the real world, it has been introduced *into* the real world. For the patient, the fictional impostor and any narrative constructed to support it, is accepted as part of her real circumstances, not as an alternative to it. The third component concerns the disjunction between real and fictional worlds. Normally the disjunction between the two prevents the possibility of action. It is impossible to interact physically

with the characters in a novel, film or play, or with those created in a private fantasy.[54] Further, setting aside fictional worlds that are designed to be interactive, the disjunction prevents us from influencing the course of events in the narrative. But because the Capgras delusion is played out in the real world rather than on a stage, in a film, or in novel, the disjunction that normally places restraints on action collapses. Of course the Capgras patient cannot interact with the fictional impostor. She interacts with her loved-one, but she does so in the delusional belief that her loved-one is an impostor. Similarly, strong action, such as killing the 'impostor', does not alter the course of the fictional narrative because one cannot physically kill a fictional character. Attempting to do so might be likened to leaping onto the stage and locking the actress in a cupboard to prevent Ophelia from drowning.

> One person can save another only if they "live in the same world'. That is, it can be true that A saves B only if it is true that A and B both exist, and it can be fictional that A saves B if it is fictional that they both exist. *Cross-world* saving is ruled out, and for similar reasons so is cross-world killing, congratulating, handshaking, and so forth. (Walton: 178, 17)

The patient only acts as she does because she believes the impostor exists in her world, that is, that they both exist in the same world.

An important difference between our interaction with a fictional character on stage and the Capgras patient's interaction with the fictional impostor is that, when at the theatre, we consciously enter a game of make-believe in which we pretend the characters are real. However absorbed in the pretence we might become, we would not claim that King Lear, for example, was a real person with whom we could interact. Likewise, actors who become totally absorbed in their roles do not behave violently towards each other. For example, the actor playing the hero does not kill the actor playing the role of villain, to save the one playing the heroine. The Capgras patient does not choose to participate in a game of make-believe because the shift to an alternative reality is the result of an involuntary, subpersonal process. The confusion between reality and fiction

[54] It is possible to interact physically with the *actors* in a play, but not with the characters they represent. We can only pretend to interact with fictional characters.

(cross-world confusion) that results, how it occurs, and why it is not corrected will be discussed in detail in the next chapter.

Following Gallagher, in the next chapter I will argue that moving between realities involves an existential shift from *being-in-the-world* of reality to *being-in-the-world* of an alternative reality. According to my account, the patient shifts to an alternative reality when viewing her loved one, and back to reality when no longer viewing her loved-one. Even if, as I will argue, the shifting between realities that underlies the delusion, is an involuntary, sub-personal process, one might suggest that the patient would be aware that it was occurring because the different existential state are phenomenologically different. Imagining winning the lottery is experientially different to actually winning it. However, this concern rests on the assumption that the non-delusional do not confuse reality and fantasy to any significant degree. The following experiment demonstrates that this is far from being true.

The Stanford Prison Experiment (SPE) was conducted in 1971 by a team of researchers from Stanford University, led by psychology professor Phillip Zimbardo. Twenty-four male students played the roles of prisoners and guards in a purpose-built mock prison on the university campus. The students were carefully screened for physical health and psychological stability. Although the experiment was to run for two weeks, it was terminated on the sixth day.

> The projected two-week study had to be prematurely terminated when it became apparent that many of the 'prisoners' were in serious distress and many of the 'guards' were behaving in ways which brutalized and degraded their fellow subjects. In addition, the emerging reality of this role-playing situation was sufficiently compelling to influence virtually all those who operated within it to behave in ways appropriate to *its* demand characteristics, but inappropriate to their usual life roles and values; this included the research staff, faculty observers, a priest, lawyer, ex-convict, and relatives and friends of the subjects who visited the prison on several occasions. (Zimbardo: 1973, 243)

By day 5, five of the previously healthy students playing the role of prisoners had to be released due to extreme stress, acute emotional disturbance and pathological behaviour. Most of those who remained adopted "a zombielike attitude and posture, totally obedient to escalating guard demands" (Zimbardo: 2007, 3). The escalating demands included brutalization, such as forcing

'prisoners' to play humiliating sexual games at night when the 'guards' thought they were not being observed. Zimbardo concluded that both prisoners and guards had become too grossly absorbed in their roles. The students being abused stayed in the experiment because they had internalized the 'prisoner' identity and came to believe they really were prisoners. They forgot they were playing a game and that the 'prisoners' were free to leave. I suggest that this was, in part, because *"from the inside* the fictional worlds ... look very much as though they are real" (Walton: 1978, 21), and there were few cues available to remind the participants it was just a game.

According to Gallagher, when entering an alternative reality, we adopt a set of beliefs and values that are not commensurate with everyday reality. This was certainly true of most of the students playing the role of guard. It was also true of the research team and faculty observers who failed to protect the students playing 'prisoner' from the abusive treatment that was causing real (not make-believe) emotional and psychological trauma.

> In many instances during our study, the participants' behaviour (and our own) directly contravened personal value systems and deviated dramatically from past records of conduct. (Haney & Zimbardo: 1998, 719)

Certainly the fact that some of the players were isolated in a mock prison for the duration of the experiment made it easier for them to detach from their real circumstances. But this was not true for all participants. The guards worked shifts, changing out of uniform at the end of shift and presumably, although not stated, returning home. The research team and faculty observers, not only went home, but also continued to work on other projects at the university. Each day these participants shifted from *being-in-the-world* of their real circumstances to *being-in-the world* of the fiction without noticing they were making, or failing to make the appropriate shift between realities. Zimbardo believed he was faithfully fulfilling his obligations as leader of the experiment when, in fact, he was still absorbed in his role as superintendent of the prison, expressing the values of that role at the expense of his ethical obligation towards his students. On the fifth day an associate professor from another university was invited to observe the experiment. She had no previous connection to the experiment, and

thus had not become caught up in the game. On seeing the mock prison and the state of the prisoners, she was appalled. She made it clear to Zimbardo that *real* people were suffering, and he recalls "that powerful jolt of reality snapped me back to my senses" (Zimbardo: 2007, 3). The experiment was terminated the next day.

The SPE involved a group of players with all the pressures and power-plays of the group supporting individuals in maintaining their fictional roles. Those surrounding the Capgras patient do not support the fiction, but rather, strive to convince the patient of her real circumstances. However, the SPE participants had the advantage of having consciously agreed to play a specific game. They knew what they were doing and why they were doing it. Although they appeared to have internalized the generating principle, rules and their roles as the game progressed, there remained the possibility of recalling, or being reminded of, that agreement. There is no such possibility for the Capgras patient because she does not consciously agree to play a game. In addition, there are no cues indicating the impostor is fictional—no book, no cinema, no stage, audience, or the like. But the most important difference between the SPE participants and the Capgras patient is that the patient has not created an alternative reality that stands against everyday reality, but rather, has introduced a fictional character into reality. As the 'impostor' concept is the meaning component of the immediate perceptual experience, the fictional character becomes part of the patient's perception of the real world. Therefore it is not possible to 'jolt' or 'snap' the patient out of the alternative reality and bring her back to her senses, as it was with Zimbardo.

Games played by the non-delusional in the real world, demonstrate how easily fiction and reality can become confused, despite the participants having consciously agreed to play a game of make-believe. This being so, the Capgras patient's confusion is less puzzling when we consider she lacks both cues and a conscious agreement that could remind her that the impostor is fictional. Perhaps the closest likeness to the Capgras experience might be that of a young

child who meets Mickey Mouse at Disneyland. Although the child knows that Mickey Mouse is a make-believe character, when interacting with the Mickey Mouse impersonator, she may accept him as real.[55]

In summary, the alternative-realities hypothesis explains the passivity exhibited by many Capgras patients, by suggesting that sourcing the 'impostor' attribution from a fictional world links it to non-action. More aggressive responses occur when the patient becomes more absorbed in, and more emotionally engaged with, the 'impostor' fiction and any supporting narrative constructed around it. Because the fiction has been introduced into the real world, the constraints on action have collapsed, there are no cues indicating the impostor is fictional, and there is no conscious agreement to a game of pretence the patient might recall.

3. Circumscription and double-bookkeeping

Monothematic delusions are described as being circumscribed because the content of the delusional belief does not appear to impact upon, or be influenced by, the beliefs held in the patient's general belief web. For example, Capgras patients do not misidentify their loved-ones when speaking to them on the phone, and they will converse with them as if everything is in order. They fail to mention the impostor. Such behaviour suggests that the content of the delusion reflects a sub-set of beliefs that is isolated from the general belief web. The existence of a separate belief web is one way to explain double-bookkeeping.

It is generally accepted that to hold a belief that is inconsistent with one's other beliefs is to be irrational. Wittgenstein says of beliefs, "What we believe is not a single proposition, it is a whole system of propositions", and "What stands fast does so, not because it is intrinsically obvious or convincing; it is rather held fast by what lies around it," (Wittgenstein: 1972, 21e; ∫141, ∫144). According to Davidson, beliefs are never irrational in themselves, but only as part of a larger

[55] This example was suggested by my supervisor, Peter Menzies.

pattern, because "nothing can count as a reason for a belief except another belief" (Davidson: 1986a, 315).

> No factual belief *by itself,* no matter how egregious it seems to others, can be held to be irrational. It is only when beliefs are inconsistent with other beliefs according to principles held by the agent himself – in other words, only when there is an inner inconsistency – that there is a clear case of irrationality. (Davidson: 1985, 348)

Characteristic of the Capgras delusion and monothematic delusions in general, is the patient's ability firmly to hold a belief, and often an additional confabulated set of supporting beliefs, that do not cohere with their general web of beliefs. Patients appear impervious to the inner inconsistency of their beliefs and to the irrationality it implies. Maher's explanationist account is inadequate because it does not address this characteristic. For example, the woman who thought she had bees in her head should have rejected it as a belief because it was inconsistent with what she already believed about bees and people's heads. Maher argues that a delusional belief is a rational explanation of an anomalous experience. But the 'bees' hypothesis should have been rejected as the explanation for the anomalous experience on the grounds that the internal inconsistency it created deemed it irrational to do so.

The alternative-reality hypothesis provides a good explanation for the circumscribed nature of monothematic delusions. Fictional worlds are created by forming a sub-web of fictional beliefs. So long as the fictional belief coheres with the other fictional beliefs in the sub-set, it is coherent. A belief only has to comply with the logical possibilities, rules, values and principles of the world to which it belongs. In terms of rationality, how it relates to belief(s) or fictional belief(s) in other worlds is not relevant. Consequently, one might rationally hold two contradictory beliefs if they belong to two different worlds. For example, one might believe that little boys cannot fly on broomsticks whilst at the same time believing a little boy named Harry Potter can do just that. This is because the second belief, as it is usually stated, is an abbreviated version of 'I believe that, in a fictional story, Harry Potter can fly on a broomstick'. When referring to shared fictions, the fictional belief is rarely stated in full, thus giving the impression the two beliefs belong to the same world. There might well be a fictional narrative in which bees live in people's heads. Certainly there are

fictional worlds in which impostors enter people's lives. If the Capgras patient's impostor is fictional, the belief only has to cohere with the other fictional beliefs in the fictional world to which the impostor belongs. But the problem for the patient is that attributing meaning in the immediate perceptual experience is a sub-personal process. Therefore she is not consciously aware of the impostor's fictional nature. When the patient says 'that man is an impostor', it is not an abbreviation of 'I believe that in a fictional world that man is an impostor'.

At a glance, the above explanation of the circumscribed nature of monothematic delusions could appear to be equally well explained by Davidsonian fragmentation. In explaining self-deception, Davidson (1986b; 1998) postulates that the cognitive system of the person in question is fragmented into partitioned compartments, between which there is no communication for the duration of the deception. Beliefs that are inconsistent with the belief at the core of the self-deception are thus, cognitively isolated from it. The problem with using this hypothesis to explain delusions is that it is essential to explaining self-deception that the lack of communication between the separated groups of beliefs be *complete*. As Klee (2004) notes, even a minimal awareness would seem to imply conscious awareness of one's inconsistency, something self-deception could not accommodate. But according to Sass, many deluded persons seem to possess some level of awareness that their delusional beliefs are not in keeping with everyday reality. Patients "often seem to have a surprising, and rather disconcerting, kind of insight into their own condition" (Sass: 1994, 3). Having described a patient of Bleuler's[56] who made fun of his own beliefs, Sass makes the following comment.

> At times, in fact, one may even begin to suspect, in the presence of such patients, that they are somehow only playacting – as if, like the protagonist of Pirandello's *Henry IV,* they are just feigning madness while taking a perverse delight in forcing those around them to support the charade. (Sass: 1994, 4)

I know of no clinical data that suggests Capgras patients have the same level of awareness as the schizophrenic patients Sass describes. However, Capgras patients who are cognizant of the bizarre nature of their beliefs and of the

[56] Bleuler (1857-1939) was a Swiss psychiatrist.

difficulty others might have in accepting them, demonstrate some level of awareness that their experience is out of keeping with everyday reality. The pervasive feeling Capgras patients report, that everything is somehow unreal or unfamiliar, also points to their awareness that all is not as it should be. The minimal awareness Capgras patients appear to have is sufficient to make the fragmentation hypothesis untenable.

Circumscription is implicated in double-bookkeeping because to act contrary to one's firmly held belief implies that belief and action are linked to two different belief webs. Double-bookkeeping is considered a characteristic of the Capgras delusion and other delusions because passive patients have a tendency to act in contradiction of their belief.

> Their failure to act on the basis of these experiences (patients who claim to be surrounded by poisoners, for instance, may nevertheless sit down happily to lunch), or to show even the slightest interest in potential corroboration or disconfirmation (which according to Jaspers are distinctive features of true schizophrenic delusions) suggests that this is a domain where reality-testing simply plays no role, a domain to which neither belief nor disbelief of the usual kind is of any relevance. (Sass: 1992, 115)

If we accept, for the sake of argument, that the meaning attribution generated by the immediate perceptual experience involves a fictional impostor, it can be further argued that although a patient may fail to distinguish between the real and the fictional at the conscious level, it does not automatically follow that the sub-personal processes also fail to make a distinction[57]. We might say the brain knows the truth of the matter even if the person does not. What I propose is that, in the case of double-bookkeeping, the patient's behaviour is driven from the sub-personal level where the distinction is acknowledged, whilst the false belief is firmly held at the conscious level, where the distinction is not acknowledged. Why this occurs is the subject of the following chapter.

[57] As will be discussed in the next chapter, distinguishing between different sources of information is a critical cognitive function. I will argue that a conflict monitor registers any discrepancy between the source of information and the world to which it is applied, and this provides an implicit awareness of the patient's actual circumstances.

4. Bizarre content

As noted in the previous section, beliefs do not exist in isolation. They exist in a supporting network of other beliefs. Reasons given for holding a belief are other beliefs from the supporting network. To firmly hold a belief whilst being unable to give any reason for doing so is considered irrational. Patients who frankly admit to the inexplicability of their belief, yet still assert its truth, do not conform to the norms of rationality. But other patients avoid the problem by constructing an elaborate system of buttressing beliefs to support the core delusional belief. Such confabulations range from the mundane to the bizarre. If the initial delusional belief is the result of a meaning attribution sourced from a fictional world, then arguably, the buttressing beliefs are constructed in accordance with the logic of the world to which the initial delusional belief belongs. For example, the belief one's family has been replaced by impostors may lead a patient to support the belief with fictional beliefs about the family having been shot by the hospital staff. If the Capgras patient's 'impostor' is drawn from science fiction she may confirm the 'impostor' belief by also believing the impostor is a robot, and therefore, could not possibly be the loved-one. The confabulations supporting the initial fictional belief might seem bizarre in terms of everyday reality, but they are not particularly strange when compared with other forms of fiction.

5. Incorrigibility

The incorrigibility of monothematic delusions is generally accepted as being quite well explained by an endorsement account. If the delusion is an experience taken as veridical, the belief is formed non-reflectively. It is suggested that as logic is not involved in its formation, the belief is not susceptible to revision in the face of rational argument. But beliefs that result from taking an experience as veridical can be revised. For example, although we perceive the lines in the Müller-Lyer illusion as unequal, we believe they are the same length. Consequently a second factor, a failure to inhibit the pre-potent doxastic response, is proposed to explain why patients do not revise their delusional belief. In my account, the second factor is the automatic stimulation of the

mechanism that enables entry into an alternative reality. Fictional realities are formed by a sub-set of fictional beliefs, that is, beliefs that are true in a fictional world. As concepts held in memory are not held in isolation, but rather, are linked to a web of beliefs relevant to that concept, the concept of a particular fictional character will be linked to information regarding that character and the world to which she belongs. The rules, values and logical possibilities of a particular fictional world are implicit in the beliefs relating to that world. Thus, we can say 'Harry Potter casts spells' because casting spells is permissible in the fictional world in which he exists. If the Capgras patient's impostor is a fictional impostor, and the second factor results in the patient entering the fictional world to which her impostor belongs, the rules, values and logical possibilities will be those that operate in that particular fictional reality. Those of everyday reality simply will not apply. Anyone attempting to point out the irrationality of the delusional belief(s) will be talking at cross-purposes with the patient. Using everyday logic to argue that the loved-one is not, for example, a Martian is like using such logic to argue that Harry Potter is not a wizard.

6. Incomplete worlds and the unexplained

Walton says possible worlds are of necessity both possible and complete, whereas fictional worlds are sometime impossible and usually incomplete. Delusional worlds that involve Martians, robots and clones clearly involve physically impossible worlds. In addition, I suggest all monothematic delusions exhibit the kind of incompleteness found in fictional worlds. Fictional worlds are incomplete because there is much about the characters and events in fictional worlds that we do not know. Only information relevant to the story is included, and unlike the real world, what is not in the story does not exist. In most fiction, the relevant information is presented little by little to maintain observers' interest in the unfolding story. Not knowing why certain events are occurring, what will happen next, who people are, their motivation or intentions, and being intrigued, puzzled or surprised, coupled with the expectation that all will be revealed in the end, is in part, what makes engaging in a story enjoyable. Knowing how the story ends can seriously reduce the

enjoyment of some stories.[58] When participating in a fictional story, lacking information is usually important, and it requires a tolerance for the unexplained that we do not normally exhibit in everyday life. There are obvious reasons why this is so. As the story is fictional, the consequences of the unexplained, however dire, are not real, and, however puzzling the situation, we expect a full explanation to be given by the story's end. We are not in danger, and we do not have to work things out for ourselves.

The contents of most delusions leave much unexplained. For example, the impostor's identity is often unknown. Who she is, where she came from and why she came, is not part of the story told. A loved-one dies in a car accident and is replaced by an impostor without the connection between the two being explained. In an unlikely act, hospital staff take a family outside and shoot them. It is claimed that systematically shooting people is something the staff do regularly, yet despite the shootings taking place in public (out front of the hospital) it appears the police have not become involved and the staff are left free to pursue this gruesome activity. Tolerance for such inconsistencies and missing information is characteristic of delusions. Whilst some patients find the inexplicable presence of an impostor distressing, others, such as the man who made his wife's impostor his mistress (Antérion *et al.*: 2008), find ways to accommodate the inexplicable. One way of accounting for the tolerance of inconsistencies, missing information and the unexplained, is to suggest that the delusional content 'impostor' is drawn from an alternative, fictional world, and that the patients' attitude towards the unexplained and the inconsistencies surrounding it, is a normal response to a character belonging to a fictional world in which such gaps in information are expected.

[58] At the end of each performance of the murder mystery play "The Mouse Trap" by Agatha Christie, the members of the audience are asked to promise they will not disclose to others how the story ends. The play has been performed continuously since 1952.

7. Conclusion

I have described how a wide range of characteristics exhibited by Capgras patients can be explained by an alternative-realities hypothesis. Using the Capgras delusion as my example, I have explained how the characteristic might arise if the meaning content of the immediate perceptual experience is drawn from a fictional world and the patient engages in the fiction. I have indicated that a second factor causes the patient to enter the alternative reality involuntarily. In the next chapter I will give a detailed description of this second factor.

CHAPTER 10: THE CAPGRAS DELUSION
A coherence-monitoring account of factor 2

1. Introduction

In previous chapters I argued that the meaning component of the immediate perceptual experience in the Capgras delusion is drawn from a fictional world, and that the 'impostor' concept, which resolves the stranger/loved-one conflict, incorporates contextual information and the patient's pre-reflective understanding that agents act for reasons. I further argued that the fictional nature of the 'impostor' helps explain some puzzling characteristics of the delusion. However, the fact that the content of the experience is uploaded as a belief without being challenged suggests a second factor is necessary for the formation and maintenance of the delusion. In this chapter I will argue that the second factor is the automatic decoupling of a conflict monitor from executive control that occurs when a cross-world conflict is detected, but cannot be resolved. According to my account, the second factor is part of the rapid, automatic, pre-reflective processing that produces the immediate perceptual experience. I will argue that the decoupling process involved in the delusion is the normal process that enables us to enter alternative realities, and that this process is not impaired, but rather employed to resolve a conflict created by an upstream neuropsychological impairment. My account is a two-factor account, not a two-deficit account.

The second factor in delusions has been described variously as the failure of a reality testing monitor (Maher: 1974), of a putative belief evaluation system (Coltheart: 2007) and of a plausibility monitor (Langdon & Bayne: 2010). Hohwy and Rajan (2102) criticise the second factor in two-factor theory on three grounds.

> It posits a domain general deficit of reasoning competence so it predicts that patients should have widespread delusions and yet patients with monothematic delusions do not. It also predicts that delusions are constantly present, instead of being, as it seems to be the case, more dynamically shifting states. It predicts that patients but not healthy controls should develop delusions in response to unusual experiences such as perceptual illusions (or

> unusual unconscious sensory processing) but, though there is very little evidence on this, it does not seem to be the case. (Howhy & Rajan: 2012, 8)

The coherence-monitoring account I offer fleshes out the general idea that the second factor involves some kind of reality testing or plausibility monitor, whilst at the same time avoiding Hohwy and Rajan's criticisms. According to Whittlesea and Williams (2001), when interpreting our environment, our cognitive processes are continuously monitored for coherence. As a function of this monitoring system, I propose that the system registers a conflict if there is a mismatch between the reality from which a concept is sourced and the reality to which it is applied. When a concept, formed through engagement with an alternative reality, is applied to the real world, the coherence monitoring system registers a cross-world conflict, thus enabling the problem to be addressed. However, I argue, that when a conflict is generated by a neuropsychological impairment, the processes normally employed in conflict situations are not able to address the problem. In response, the conflict monitoring system decouples from executive control. I suggest that the decoupling process is a normal process designed to allow participation in alternative realities, and this normal process is brought to bear on an abnormal psychological event that is causing cognitive dissonance. Therefore, I propose that the second factor in the Capgras delusion is the automatic and involuntary decoupling of the conflict monitoring system from executive control.[59] In making this suggestion I draw on social cognitive and dissociated control theories of hypnosis and neuroimaging data from the investigation of cognitive conflict and control mechanisms.

2. Source monitoring

According to Whittlesea and Williams, people are chronically involved in constructing percepts and cognitions about their environment, whilst at the same time monitoring those processes.

> In this act of constructing an understanding of their environment, people also chronically evaluate the quality or coherence of their processing. This evaluation or monitoring function

[59] In Chapter 11 I will suggest how this second factor may be applied to several other monothematic delusions and how it might be implicated in schizophrenia.

> enables them to detect errors or inconsistencies in their processing. (Whittlesea & Williams: 2001, 14)

The process of constructing percepts and cognitions about our environment involves accessing information from memory. Information held in memory is derived from various sources and needs to be applied in keeping with its source.

> Because humans have a cognitive system that takes information from a number of perceptual sources and that can itself internally generate information as well, one of the mind's most critical cognitive functions is discriminating the origin of information. (Johnson: 1991, 180)

Johnson, Hashtruodi and Lindsay (1993) propose *source monitoring* as a general term for the monitoring of the origins of such information. Researchers have proposed that the source of our memories is subject to monitoring to enable us to differentiate between memories of things and events that were real and those which were dreamed, imagined or part of a fictional narrative. Failure to attribute the correct source of information in memory retrieval can result in false memories (e.g. Dodhia & Metcalfe: 1999) and confabulation (e.g. Johnson: 1991; Burgess & Shallice: 1996), and "source monitoring errors have been found to be more frequent among schizophrenic individuals than among healthy individuals" (Vinogradova et al.: 1997, 1530). Empirical research primarily investigates the monitoring of the origins of conscious recollections, proposing several theories that describe how this might be done. My interest is not in the general recall of memories but in the retrieval of information from memory for the attribution of meaning to the immediate perceptual experience. Therefore my interest is also in the sub-personal mechanism that enables us to distinguish between the fictional and the real and to apply concepts to their correct worlds.

3. Keeping different worlds apart

As Gallagher notes, the experiencing subject does not live in one unified world of meaning, but in sub-universes or finite provinces of meaning. As well as the shared reality of everyday life, there are alternative realities that take us away from the everyday world. These are the realities of dreams, daydreams, imaginings and all the many forms of fiction in which a subject might engage. In their normal, everyday activity, the non-deluded have little difficulty keeping

the different worlds apart. As early as age two, children can differentiate between imagination and reality (e.g. DiLalla & Watson: 1988; Wellman & Estes: 1986; Woolley& Phelps: 1994), and research suggests that by four years they are capable of keeping multiple pretend game worlds apart (Weisberg & Bloom: 2009). By adulthood a fully developed sub-personal mechanism keeps track of the different worlds with their different values, truths and logical possibilities, and prevents subjects from becoming lost in a confusion of multiple realities. To avoid cross-world confusion it is necessary to differentiate between alternative realities in three different ways. The first identifies the reality in which we are engaged. (Are we in the real world or the world of a novel, daydream or other fiction?) The second identifies the source of our conscious memories. (Did that event really happen or was it something I imagined?) The third identifies the nature of the world to which a concept belongs. (Is a unicorn a real animal or a fictional beast?) For all three, the distinction is usually made pre-reflectively at the time of the immediate experience. But the distinction made in the immediate moment can be challenged and subjected to scrutiny. "I had to pinch myself to make sure I wasn't dreaming". "I remembered meeting her when I was a young child but then realized she died before I was born". "That unicorn I see in the woods cannot possibly be real". My interest is in the third distinction, our mind's ability to determine the *source* of the concepts utilized when attributing meaning to the immediate perceptual experience, and to ensure their application is appropriate. I suggest it is an anomaly in this system that constitutes the second factor in the Capgras delusion.

The sub-personal cognitive process that determines the source of a concept is a matter of conjecture. But one way I suggest the mind might do so is by recalling the effects of cognitive processes and mental states that accompanied the perceptual event(s) giving rise to the concept.

> Arguably, memory experience can include a greater range of high-level content, since what is retrieved and recollected is not (merely) the perceptual features of an encoded event, but those perceptual features plus the effects of cognitive processes, which were engaged at the time of encoding and which also come into play at the time of the retrieval (Conway, 1996), and which embellish and filter (see, e.g., Mitchell & Johnson, 2000). (Langdon & Bayne: 2010; 339)

Langdon and Bayne are discussing memory experience. But their comments could be applied, in principle, to any information retrieved from memory, and that includes the concepts retrieved from memory to provide the meaning content of the immediate perceptual experience. If Gallagher is correct and it requires an existential shift to *be-in-the-world* of a fiction, concepts drawn from a particular fiction may be accompanied by a trace of the existential state that enabled the subject to engage in the fiction to which the concept belongs. Tagging concepts in this way would identify them as belonging to a particular world and thus place constraints on their application. For example, if the name Ophelia or Harry Potter comes with an existential 'tag' identifying the person as belonging to a fictional world, it would enable the subject to respond to her or him appropriately, that is, as a fictional character rather than as a real person. If Capgras patients have developed the 'impostor' concept through engagement in fictional worlds, as I argued previously (see Chapter 8.), there would be a fictional character 'tag' attached to the concept providing the meaning content of the immediate perceptual experience. When the fictional character attribution is applied to the perception of a real person in the real world, cross-world confusion occurs. If it is not detected and corrected, it causes the patient to respond inappropriately. That is, whether the patient responds passively or aggressively, she is responding to the fiction rather than to the reality of the world in which her action is taking place. The incoherence caused by cross-world confusion should be detected by the coherence monitor, thus enabling higher cognitive controls to override the error. I propose that in the case of the Capgras delusion, the higher cognitive controls are not able to override the error because the conflict monitor and executive control are automatically and involuntarily decoupled for the following reason. Maher notes, that when checking an anomalous perception with an underlying biological cause involves taking a closer look at the environment, "it will serve to validate rather than refute the experience of the patient" (Maher: 1974, 108). In the case of the Capgras delusion, the upstream neuropsychological impairment causes the fiction to be generated every time the patient views her loved-one. Therefore, checking merely confirms the initial perception that the loved-one is an

impostor. I suggest it is the persistence of the cross-world conflict and the cognitive dissonance it causes, that stimulates the decoupling process. Once this has occurred, offline resources become unavailable, and the problem cannot be resolved reflectively.[60]

Tagging the meaning attributions according to their existential status may also serve another purpose. Whittlesea and Williams (2000; 2001) hypothesise that when there is an internal inconsistency in face processing, one that violates expectations, it creates a *phenomenal response*, that is, a feeling of familiarity. The feeling of familiarity alerts the subject to the presence of an inconsistency, enabling her to consciously seek a resolution. When the person is identified the feeling abates. Similarly, the incoherence created by cross-world confusion could create a phenomenal response. In the Capgras delusion, this might be the phenomenal response patients describe as a feeling that everything is unreal or strange in some way. The feeling, that reality is not quite how it should be, could be an expression of the fictional (unreal) origin of the concept. However, later I will argue that when the decoupling process is stimulated in the Capgras delusion, the patient automatically and involuntarily enters the alternative reality of fiction to which the 'impostor' belongs. Thus, the phenomenal response may result from the conflict between two co-existent existential states, that of the real, and that of fiction.

Various 'monitoring' explanations of the second factor in monothematic delusions describe the problem as a failure of a monitoring system to detect the implausibility of a thought or belief in relation to reality. Yet there has been little investigation of the putative failure of the monitoring system, leaving the explanations saying little more than that patients fail to detect the irrationality of their thoughts and beliefs. If the problem is described more specifically as the

[60] Checking the environment more closely also does not resolve illusions, such as the Müller-Lyer illusion. The lines continue to be perceived as being of unequal length. The difference between this and the Capgras delusion is that the illusion does not involve cross-world conflict. The misperceived lines are real lines in the real world.

failure to monitor or respond to a conflict between data (e.g. wife/stranger), or belief webs (e.g. about bees and people's heads), or between fictional and non-fictional worlds, or between other cognitive processes, the available data on conflict monitoring could be brought to bear on the problem. Neuroimaging has been, and is being, used to isolate the brain areas involved in conflict monitoring both in normal subjects and those in an induced hypnotic state. Hypnosis is relevant to delusions because "hypnotic analogues can produce compelling delusions with features that are strikingly similar to their clinical counterparts" (Cox & Barnier: 2010, 202). Beliefs generated by the patient's experience whilst in the hypnotic condition are held with strong conviction, maintained in the face of overwhelming evidence to the contrary and are resistant to counterargument.

> The distinguishing feature of hypnosis appears to be the subjective state; and the main feature of this state is the hypnotized subject's emotional conviction that the world is as suggested by the hypnotist, rather than a pseudoperception of the suggested world. (Sutcliffe: 1960, 200)

According to Sutcliffe, hypnotised subjects essentially become deluded about the actual state of the world. Further, "in addition to modelling the features of clinical delusions, hypnosis may also be able to model some of the underlying processes" (Cox & Barnier: 2010, 209). The similarities between delusions and the beliefs held by hypnotised subjects suggests the likelihood of there being a common underlying mechanism, and, therefore, that research into the cognitive mechanics of hypnosis is relevant to research into the causes of delusion. It also suggests that the neuroimaging data on conflict monitoring in the hypnotic condition might provide insight into the same mechanism in delusion. Therefore, before discussing the neuroimaging data, the next section will be devoted to a discussion of hypnosis and delusions.

4. Hypnosis and delusions

Experimental hypnotic analogues are used to recreate the features of clinical syndromes in the laboratory and to model the hypothesised processes that contribute to clinical disorders. Hypnosis is considered a suitable tool for investigating delusions because of the similarity between their features. These

include the phenomenological features of the experience, the conviction with which beliefs are held, and their resistance to counterargument and overwhelming evidence to the contrary. Examples of monothematic delusions that have been induced using hypnosis are reverse intermetamorphosis[61] (Cox & Barnier: 2009a; 2009b), mirrored-self misidentification (Barnier *et al.*: 2008), somatoparaphrenia (Rahmanovic et al: 2012), alien control (Blakemore, Oakley & Frith: 2003) sex change delusion (McConkey, Szeps & Barnier: 2001), erotomania (Attewell et al.: 2012) and folie á deux (Cox & Freeman[62])

The plethora of theories purporting to explain the mechanism underlying hypnosis is not the subject of this thesis, therefore, I will restrict my discussion to two influential theories, dissociation theory and social cognitive theory. Both theories cover a wide range of views but briefly, dissociation theorists hold that the hypnotic state is caused by some form of dissociation in cognitive processes, whilst social cognitive theorists emphasise hypnosis as a social interaction involving motivation, expectancies and goal orientation. The dissociation and socio-cognitive theories have been regarded as alternative, conflicting theories, but recent information from research "make possible a rapprochement between theoretical accounts that have vied for attention and empirical support" (Lynn & Green: 2011, 277). This rapprochement is evident in Cox's claim that "hypnotic delusions arise due to a combination of cognitive (e.g. dissociative) and social/motivational processes" (Cox & Barnier: 2010, 225). In the following discussion I accept the general assumption that the strong similarity between the features of hypnosis and delusions indicate the likelihood of a common underlying mechanism. Of particular interest in relation to my proposed second factor in the Capgras delusions are the use of imagination (fictional worlds) to

[61] Intermetamorphosis is a condition in which the patient sees others changing into someone else both in external appearance and personality. Reverse intermetamorphosis is a condition in which the patient sees the changes occurring to themself.

[62] This successful hypnotic induction of folie á deux was conducted and videotaped by Rochelle Cox and Luke Freeman on May 24th 2012. The videotape was viewed by members of the CCD Belief Group at Macquarie University on May 25th 2012. An account has not yet been submitted for publication.

induce the hypnotic state, the social context in which it takes place, and the putative cognitive dissociation that enables the hypnotic state to occur.

i) Imagination and social context

Imagination is an integral component of the hypnotic experience. "Imaginative involvement plays a central role both as background for hypnotic susceptibility and as a feature of hypnotic behaviour" (Hilgard, J.: 1974, 138). Kihlstrom's widely accepted definition of hypnosis describes it as

> a process in which one person, designated the hypnotist, offers suggestions to another person, designated the subject, for imaginative experiences entailing alterations in perception, memory and action. In the classic case, these experiences are associated with a degree of subjective conviction bordering on delusion, and experienced involuntariness bordering on compulsion. As such, the phenomena of hypnosis reflect alterations in consciousness that take place in the context of social interaction. (Kihlstrom: 2008, 21)

Hypnosis could be viewed as a social activity in which two persons agree to play a game of make-believe. Before the game begins, the role each participant will play is clarified. Using Walton's terminology, the generating principle would be something like 'when person A suggests an imaginative experience, person B imagines having that experience'. Games of make-believe involve *being-in-the-world* of fiction, perhaps reflecting an alteration in consciousness (Kihlstrom), or an altered existential state (Gallagher). When engaging in fictional realities, Walton stresses the importance of our ability to experience the fictional characters and events as somehow being real, even though we know they are not. Similarly, successful hypnosis does not merely require the subject to *imagine* the suggested experience, but to *have* the experience and experience the suggestion as *real*. The following instruction is taken from the flow sheet for establishing rapport in the Stanford Hypnotic Susceptibility Scale Form C.

> The idea is just to pay attention to what is going on, and try to have the experiences that are suggested to you. (Weistzenhoffer & HiIgard, E.: 1962, 3)

And using an adaptation of the Stanford Hypnotic Susceptibility Scale Form A (Weitzenhoffer & Hilgard: 1959), Braffman and Kirsch offer the following introduction to the subject. "In this study, we want to assess your ability to use your imagination (hypnosis) to experience various things that will be described to you on audiotape" (Braffman & Kirsch: 1999, 579). Again, two instructions

taken from the standard hypnotic induction procedure used by researchers at Macquarie University are "Listen carefully to my voice, and the things I'm going to say to you, and to the things I'll ask you to experience" and "You are going to experience many things that I will ask you to experience" (MACCS hypnosis flow sheet)[63]. So hypnosis might be better described as a game of make-believe in which one of the two participants accepts an imaginative suggestion from the other and experiences it as real. For example, the hypnotised participant does not just *imagine* her arm is moving of its own accord, but *experiences* it as doing so.

My hypothesis is that the same mechanism enables subjects to enter alternative realities of all kinds, including the hypnotic condition and some delusions. Further, the degree to which subjects are able to become absorbed in an alternative reality generally, will determine their susceptibility to hypnosis, and, given a first factor, to delusion. A finding from early investigations of hypnosis was that imaginative involvements outside of hypnosis, often originating in early childhood, were related to hypnotic susceptibility.

> These involvements included reading, the dramatic arts, religion, some forms of sensory stimulation, imagery and imagination (including imaginary companions), adventuresomeness, and some aspects of sports, especially the less competitive ones. While one or two strong involvements often sufficed as pathways to hypnosis, the very highly susceptible commonly had a number of areas in which they could become deeply involved, temporarily setting reality aside as they savoured the experience. (Hilgard: 1974, 138-139)

I predict that further investigation would similarly uncover a relationship between such imaginative involvements and susceptibility to developing delusions.

ii) Dissociation

Using dissociation to explain hypnosis was originally proposed by Janet (1901), who suggested that a subunit of mental life becomes split off from the rest, and separated by amnesic barriers from both awareness and voluntary control. This creates a constriction of normal awareness of some material, enabling a special

[63] Flow sheets for proposed creation of delusions under hypnosis are provided to members of the CCD Belief Group who meet regularly at Macquarie University.

influence on behaviour of this material through activation by suggestion. However, research results did not confirm Janet's hypothesis, and his description of amnesic barriers was considered implausible. Ernest Hilgard (1977) proposed a neodissociation theory, focusing on conscious versus unconscious processes in dissociation. He suggested there is a splitting off of certain mental processes from the main body of consciousness, with various degrees of autonomy. In his theory, parallels streams of consciousness co-exist, separated by amnesic barriers. He tries to fuse this with Janet's theory but tends to "mix together incomplete explanations across somewhat vaguely defined levels" (Woody & Sadler: 2008, 84), and his "cognitive control model of hypnosis remained frustratingly incomplete and vague" (Jamieson & Woody: 2008, 113). Nevertheless, an important speculation Hilgard makes is that the dissociation of processes concerns the possible loss of integration of the monitoring and control functions of the executive system. In response to the implausibility of the amnesic barriers, Bowers (1990) reformulated the neodissociation theory. He noted that one of the confusions in Hilgard's account concerns the conflation of dissociated experience and dissociated control, and he proposed that the theory should be split in two distinct sub-theories. Later he focuses on the dissociation of control, suggesting that in hypnosis, lower subsystems of control are activated more or less directly by suggestion, bypassing the executive level, along with the processes of volition and sustained effort. "Bowers believed that hypnosis alters the control of behaviour, rather than distorting the perception of this control" (Sadler & Woody: 2010, 155). Following Bowers, many different versions of both dissociated control theory and dissociated experience theory have been proposed. Hilgard suggests that hypnosis involves a process dissociation that results in the loss of monitoring and control functions of the executive system. It is a suggestion I am pursuing although, in keeping with neuroimaging data, I propose there is dissociation *between* conflict monitoring and executive control rather than a loss of both.

5. Conflict monitoring, executive control and neuroimaging

Cognitive control mechanisms in normal subjects have been the subject of much research using neuroimaging techniques and hypnosis also has been researched using this technique. At the outset it needs to be noted that the use of neuroimaging to investigate cognitive functions is a very inexact science. "fMRI signals are weak and engulfed by much "noise" in the form of false signals" (Raz: 2011, 368), and "it takes a great deal of computer processing and human judgment to get from blood oxygen levels to a snapshot of altered consciousness" (Raz: 2011, 367). The common practice of 'reverse inference', that is, inferring a specific mental state from activation of a particular brain region, is also problematic. "It is not a logically valid form of deductive reasoning" (Poldrack & Wagner: 2004, 180). Having said that, the involvement of the anterior cingulated cortex in response to conflict in certain circumstances is "one of the most firmly established findings in all of cognitive neuroscience" (Botvinick, Cohen & Carter: 2004, 539).

A number of studies have been specifically designed to differentiate the neural substrates of conflict detection from those of cognitive control.

> Based on studies that have attempted explicitly to tease apart conflict and control processes, it appears that the fronto-parietal executive attention network can be broken down into a component that is primarily involved with detecting conflict (the dACC) and another component primarily dedicated to strategic adjustments in control (the lPFC).[64] (Egner & Raz: 2008, 34)

Interference tasks such as Stroop have provided opportunities to research the neural substrates of cognitive control processes using neuroimaging[65]. In fact "The Stroop task has evolved into perhaps the primary psychological measure of high-level 'executive' cognition" (Egner & Raz: 2008, 32). As a result of the interpretation of neuroimaging data, the following theory of conflict monitoring has been proposed.

---

[64] The dACC is the dorsal anterior cingulate cortex. The lPFC is the lateral prefrontal cortex.

[65] Stroop involves a series of tests in which participants are presented with words printed in a colour not denoted by the name, for example, the word blue printed in red ink. Based on time delay variables between tests, it has been concluded that the mind automatically determines semantic meaning and has to override this response to enable the participant to name the colour of the ink.

> The conflict monitoring theory proposes that regions within dorsal medial frontal cortex (dMFC), including the anterior cingulate cortex, monitor for the occurrence of competition, or conflict, in action selection to detect and signal the need for an increased cognitive control. (Yeung, Cohen & Botvinick: 2011, 316)

Yeung, Cohen and Botvinick describe the conflict as being related specifically to 'action selection' but other writers describe the conflict in terms of information processing more generally.

> One hypothesis concerning the human dorsal anterior cingulate cortex (ACC) is that it functions, in part, to signal the occurrence of conflicts in information processing, thereby triggering compensatory adjustments in cognitive control. (Botnivich, Cohen & Carter: 2004, 539)

And again

> We propose that the anterior cingulated cortex (ACC) contributes to cognition by detecting the presence of conflict during information processing, and to alert systems involved in top-down control to resolve this conflict. (van Veen & Carter: 2002, 478)

Van Veen and Carter claim ACC activation may not be limited to response conflict, but might respond to other sources of conflict as well. They observed ACC activation, which they interpreted as indicating conflict at a conceptual level, in participants reading stories that did not form an integrated narrative. They noted that this was "consistent with previous findings that suggest that ACC activity can be elicited in situations without motor requirements" (van Veen & Carter: 2002, 478). Taken together the above demonstrates a general consensus that the dACC is a brain area which monitors conflict in cognitive processing and engages cognitive controls in the IPFC to resolve the conflict.

> Neuroimaging studies have revealed a complex network of brain areas that is involved in cognitive control, including among others areas the PFC, ACC and PPC[66]. This vast database of findings across disciplines constrains the development of theories of cognitive control. (Davelaar: 2008, 115)

Despite the inexactness of neuroimaging in cognitive research, the results cannot be ignored.

Administering Stroop tasks to subjects in an induced hypnotic state has revealed a disturbance in the relationship between the conflict monitoring system and executive control system.

[66] The PPC is the parietal cortex, part of the network of associated brain areas investigated by Davelaar. For a description of association cortices see Bloom, Nelson & Lazerson (2001, 320) *Brain, Mind, and Behavior* 3rd edit.

> It appears that the increased conflict-related ACC activation in highly susceptible subjects in the hypnosis condition was not accompanied by a concurrent strategic increment in cognitive control, as would be expected under normal circumstances (Botvinick et al. 2001, 2004: Kerns et al. 2004). These fRMI data suggest the possibility of decoupling of conflict monitoring and cognitive control function in highly susceptible subjects after hypnotic induction, corresponding to a breakdown in the functional integration of two key components of the frontal attention system. (Egner, Jamieson & Gruzelier: 2005, 975)

The data provided by Egner, Jamieson and Gruzelier suggests that the monitoring system registers a conflict, indicated by increased conflict-related ACC activation, but that the registering of a conflict does not trigger top-down executive control to resolve the conflict, indicated by a lack of the normally expected increased activity in the LFC[67].

> The associated increase in activity in anterior cingulate cortex in the absence of compensatory changes in left frontal cortical areas has been interpreted as evidence that hypnosis acts to decouple the normal relationship between conflict monitoring and cognitive control. (Oakley & Halligan: 2009, 265)

The generally accepted interpretation of the data from neuroimaging is that it is the decoupling of conflict monitoring from executive control that enables highly susceptible subjects to enter the hypnotic condition, and consequently to accept suggested implausible experiences as real. But data also suggests that hypnosis may merely strengthen highly suggestible subjects' normal response to suggestion.

> There is good evidence, however, that subjects can respond to suggestions of this sort without the need to employ formal induction procedures. Indeed, the best predictor of the suggestibility of an individual in hypnosis is their responsiveness to the same suggestions outside hypnosis. Nevertheless, hypnotic induction procedures can increase responsiveness to suggestion, particularly if expectancy has been raised by explicitly labelling the procedure 'hypnosis'. (Oakley & Halligan: 2009, 264)

This is consistent with the idea that hypnosis is a form of imaginative game playing that is enhanced by the setting and cues provided by the hypnotic induction process.

6. Decoupling: a useful mechanism

Although the decoupling described above is a sub-personal process, hypnosis demonstrates that the process is responsive to intention. Undergoing hypnosis

[67] LFC is the lateral fontal cortex.

is a voluntary act.[68] As evolution is unlikely to have produced a mechanism purely to enable a minority of humans to enter a hypnotic state, it may be assumed that the ability to voluntarily decouple the executive control function from the conflict monitoring system plays an important role in some aspect of human cognition. I suggest that this crucial function has evolved to enable humans to enter imaginative states.

> It would be hard to exaggerate the benefits we reap from our capacity for imagination. The imagination is critical to hypothetical reasoning, planning and creativity. It is central to science and even more central to philosophy. (Nichols: 2008, 518)

It also enables us to enter the alternative reality of fictional worlds. Arguably, our default world is that of everyday reality, as those given to distancing themselves from reality by entering alternative worlds are more likely to be eaten by tigers. Staying in the real world has survival advantages. Therefore, it can be argued that the monitoring system evolved to detect any processing output that conflicts with our understanding of everyday reality. However, when it *is* safe to enter an alternative reality, we do not want to be continually distracted by an alarm system that keeps reminding us that the fictional world we are enjoying is not real. To engage in a fictional world we need to be able to temporarily suspend disbelief, and one way to do this might be to decouple the executive control from the conflict monitor.

Common experience indicates that our ability to enter fictional worlds is variable. Sometimes we are more engrossed in the fictional world than at others. The more engrossed we are, the less aware we are of everyday reality. Arguably, decoupling of the conflict monitor from executive control serves to disarm the pre-potent default response that normally keeps us anchored in everyday reality. Our experience of variable degrees of absorption in alternative worlds suggests that the decoupling can occur in degrees, with the greater the decoupling, the greater our absorption in the fiction and the more we are able to experience the fiction as real. In a review of imaging data investigating the hypnotic condition conducted by his colleagues and others, Kirsch notes "our

[68] As previously noted (Chapter 4.), Gerrans claims that one of the essential properties of imagination is that it is potentially subject to voluntary control.

data are consistent with various positions on the altered state continuum" (Kirsch: 2011, 356).The observations made by Oakley and Halligan noted in the previous section, also suggest that hypnosis is not an all-or-nothing condition, but rather the extreme of a continuum. The degree to which highly suggestible subjects respond to suggestion outside of hypnotic induction is consistent with the idea that decoupling is continuum based, as is Kirsch and Lynn's (1995) argument for the altered state continuum in hypnosis. This is in keeping with our common experience that indicates imagination is not an all or nothing state but a state that allows for degrees.

7. Capgras delusion: the second factor

My argument is that the problem underlying the Capgras delusion is a disturbance in the meaning component of the immediate perceptual experience. I have described how a delay in face encoding produces a stranger/loved-one conflict which is resolved by accessing the 'impostor' concept held in memory. However, if the 'impostor' concept was acquired during engagement in fictional worlds, I have suggested it will be accompanied by an existential tag classifying the impostor as fictional.

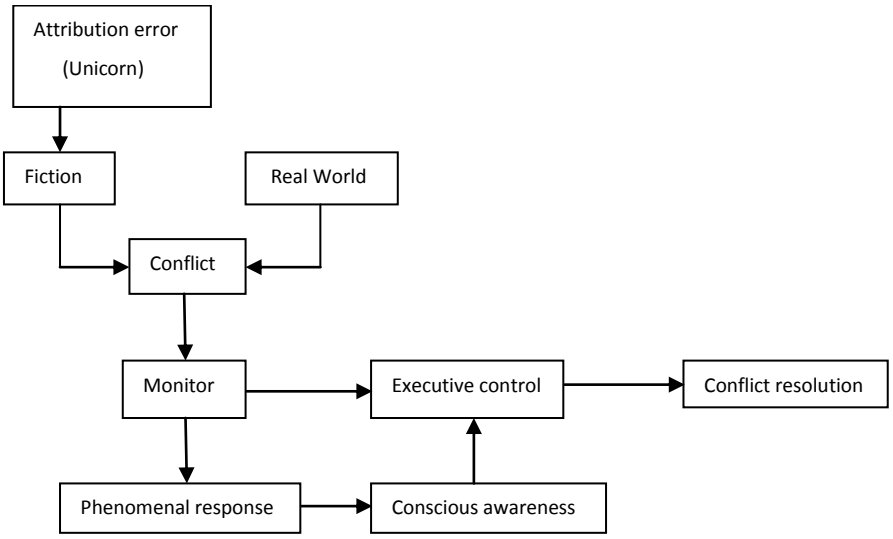

Figure 5: An attempt to insert a concept from a fictional world into the real world causes a cross-world conflict. The conflict monitor alerts executive control and stimulates a phenomenal response, which engages conscious awareness. As a result the conflict is resolved (error corrected).

When the concept is applied to the real world, cross-world confusion occurs. This creates a conflict that normally would be detected by a conflict monitor, which would trigger a response from executive control. The way the system normally works can be described using the following example. If an unusual configuration in the environment causes a subject falsely to perceive a unicorn in the woods, the perception of a fictional beast in the real world creates cross-world confusion. Activation of the conflict monitor triggers the executive control and the production of a phenomenal response, such as surprise, designed to stimulate conscious awareness. This chain of responses enables the subject to correct the processing error. As the 'unicorn' attribution is a misinterpretation of the environment, it can be corrected by examining the environment more closely. Even a conscious awareness of the conflict could lead to the dismissal of the attribution as impossible, enabling the unusual configuration to be perceived correctly (see Fig 5.). There is no such simple solution for Capgras patients because the problem is not a straight forward one-step error. The 'impostor' attribution is the cognitive solution to data conflict produced by impaired face encoding. Correcting the 'impostor' misattribution leaves unresolved data conflict upstream. In addition, although the 'impostor' attribution is a misinterpretation of the environment, it is not the result of an external anomaly (e.g. an unusual configuration of visual data) but of an *internal* anomaly. Therefore checking the environment more closely does not help.

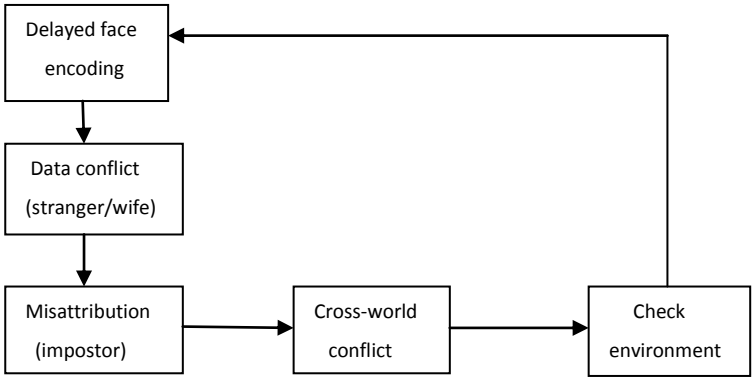

Figure 6: The vicious cycle created by the delayed face encoding in the Capgras delusion.

In fact it compounds the problem by re-stimulating factor one, thus leading to a vicious cycle. What I propose is that to break the vicious cycle and end the cognitive dissonance, the conflict monitor automatically decouples from executive control, and disbelief is suspended with regard to the impostor. As previously noted, hypnosis research demonstrates that the decoupling can be so complete that fiction can be experienced as reality. I suggest that this occurs in the case of the Capgras delusion. To remove the discomfort created by an irresolvable cross-world conflict, the decoupling in the Capgras delusion is sufficiently complete for the impostor fiction to be experienced as reality.

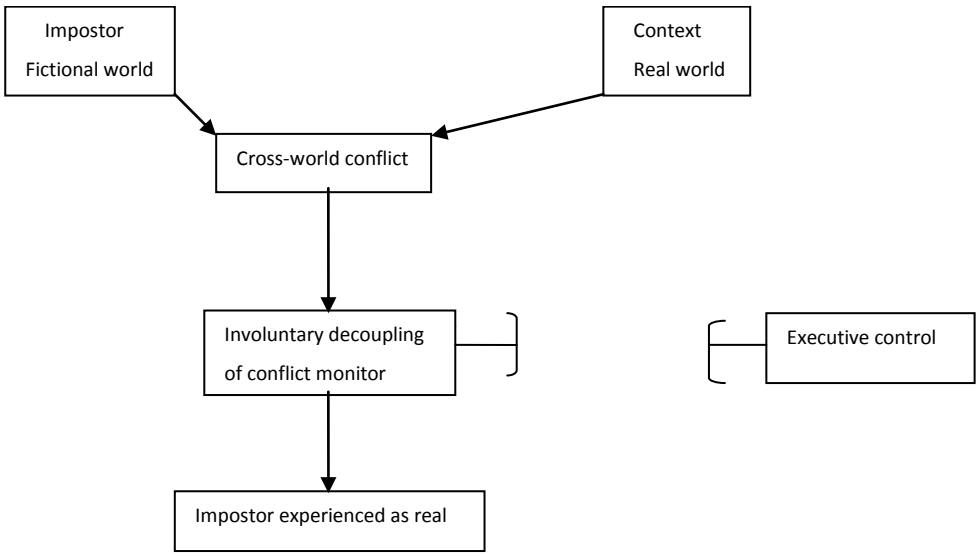

Figure 7: To avoid the vicious cycle, involuntary decoupling of the conflict monitor and executive control occurs to enable the fiction to be experienced as real, thus cohering with the experience of the everyday world as real.

Figure 7 shows how experiencing the fictional impostor as 'real' enables the attribution to cohere with the patient's experience of the real world. It needs to be emphasised that the cross-world conflict is not the second factor in the Capgras delusion. The second factor is the automatic, involuntary decoupling of the monitor from executive control, which is an attempt to resolve a conflict that cannot be resolved using the normal processes designed to resolve such conflict. The decoupling allows entry to a fictional reality, enabling the 'impostor' fiction to pass unchallenged, much in the same way as a fictional character is accommodated by members of a theatre audience. If a non-delusional person generates the impostor concept as part of the immediate perceptual experience,

and applies it to a real person in the real world, I assume that the same cross-world conflict occurs for the same reason described for the Capgras patient. The difference is that there is no neuropsychological impairment driving the perception. Unlike the Capgras patient, the 'impostor' attribution is not generated every time she looks at her loved-one. Looking more closely is unlikely to reproduce the initial experience, and, therefore, there is no reason for the decoupling process to activate.

An important difference between the Capgras delusion and the normal entering of a non-delusional alternative reality is that entering imaginative thought, the world of a representational work of art, or the hypnotic condition can be a voluntary act.[69] And just as entering these states is potentially voluntary, exiting them is also potentially voluntary. But for the Capgras patient, engaging in the fiction is not a voluntary act. The decoupling is an automatic, involuntary process, a conflict resolution strategy necessitated by the output of an upstream anomaly cause by cognitive impairment. As long as the upstream anomaly creating the known/unknown conflict continues to drive the process, the patient is forced to remain a captive of the fiction.

8. In reply to Hohwy and Rajan

Hohwy and Rajan suggest that a general deficit should result in widespread delusions. This is particularly true of Coltheart's two-deficit theory that proposes the second factor is impairment to a belief evaluation system. One might expect this to lead to the poor evaluation of a wide range of beliefs. The criticism does not apply to my hypothesis, because the second factor I propose does not involve a cognitive deficit or impairment. The decoupling I describe is a normal process brought to bear on a very specific abnormal situation. The formation of other delusions would require further cognitive impairments, resulting in data conflicts that are resolved by drawing meaning from an

[69] Sometimes a train of imaginative thoughts is triggered by an aspect of the environment rather than being intentionally generated and therefore, is not considered a voluntary act. For a discussion of the voluntary nature of imagination see Gerrans (2012).

alternative reality, thus creating cross-world confusion that triggers the decoupling. Hohwy and Rajan's second criticism is that two-factor theory predicts that delusions are constantly present when they tend to come and go. I disagree with this criticism. It would only hold true if the evaluation or monitoring system was completely destroyed. If impairment is conceived of as a weakening of the system, rather than its complete destruction, it is likely that at some times the system will fail and at other times it will not. Any factor that affects the efficiency of normal cognitive processing, such as fatigue, cognitive overload, emotional states, physical ailments, and so forth, would render a weakened system more likely to fail at certain times than at others and thus create variability in the condition. Hohwy and Rajan's third criticism is that patients should develop delusions in response to perceptual illusions or unusual sensory processing, but that there is little evidence of this. We might expect, for example, that patients with impaired belief evaluation or monitoring systems would have difficulty overriding the false belief that the lines in the Müller-Lyer illusion are uneven. According to my hypothesis, delusions result when there is a cross-world conflict, and the error cannot be corrected by checking the environment. Neither is the case with the Müller-Lyer illusion. Long and short lines exist in the real world, and the lengths of the lines can be checked with a ruler. Illusions, such as Bill and Ben in the Ames room are a little different.[70]In the hypothetical case of a delusional mother viewing her twins in the Ames room, it is unclear how she would resolve the conflict between the illusion and her prior knowledge. Rooms that alter people's heights do not exist in the real world, so the meaning attributed to the experience might be drawn from an alternative reality. On the other hand, the mother might hold the prior belief that cognitive scientists try to trick subjects volunteering for experiments. In this case, the illusion is in keeping with her prior belief and does not create cross-world confusion. Whichever the case, she could check the environment.

[70] An Ames room is constructed so that two people of the same height, standing on opposite sides of the room, appear to be of unequal height to an observer. It is suggested that in the hypothetical case of the mother with twins, Bill and Ben, the mother would reject the veracity of the visual illusion because of her prior knowledge of her sons' heights.

According to my hypothesis there may be some illusions that lead to delusion, but I suggest this would not normally be the case. As I am unclear what Hohwy and Rajan have in mind when they refer to 'unusual unconscious sensory processing' I will not attempt to address that issue.

9. Ventromedial frontal lesions

In arguing for a second factor Coltheart (2007) cites Tranel, Damasio and Damasio's (1995) study of ventromedial patients. He notes that these patients have the same deficit as Capgras patients, but they do not have the delusion. The experiments conducted in the study "were designed to provide a contrast between *overt* and *covert* discrimination of familiar faces" (Tranel, Damasio & Damasio: 1995, 425). The study showed that "ventromedial frontal subjects failed to generate discriminatory covert-level responses (SCRs) to the familiar faces they so easily recognized" (Tranel, Damasio & Damasio: 1995, 428). Coltheart argues that the presence of the deficit in the absences of the delusion demonstrates "something else is needed if this kind of failure of autonomic response is to lead to the Capgras delusion" (Coltheart: 2007, 1048). According to Coltheart, the 'something else' is a second deficit, that is, impairment to a belief evaluation system. Coltheart's account can accommodate a complete absence of delusions in ventromedial patients because his second factor is a separate deficit, which may, or may not, co-exist with the first deficit. However, some ventromedial patients *could* have the second deficit, and therefore, the delusion. In my account, the second factor is not a deficit, but the degree to which the patient becomes absorbed in a story, and the degree to which the conflict monitoring system decouples from executive control. Therefore, given a large enough sample, some ventromedial patients *should* develop the delusion.

Tranel, Damasio and Damasio do not discuss delusions and provide no data concerning the delusional or non-delusional status of the participants. The study was conducted using black-and-white slides, with a neutral background and no features below the neck. Lucchelli and Spinnler's observation that the Capgras delusion "is never reported to be triggered by looking at photographs

of the target person" (Lucchelli & Spinnler: 2008, 234), suggests that ventromedial patients would not have had a delusional response in this test situation even if they were susceptible to developing delusions. There is no indication of their delusional status outside the laboratory situation. Therefore, the non-delusional status of the patients in their study is an assumption. Further, there were only four frontal ventromedial subjects in the study. It is not possible to draw conclusion about all ventromedial subjects based on the data obtained from four test subjects. Therefore, the claim that all ventromedial subjects have the deficit but not the delusion, cannot be made based on this study. My prediction is that, given a sufficiently large sample, some ventromedial subjects would experience the Capgras delusion, although not in the laboratory. At present, the data is insufficient to resolve the issue.

10. Testing the hypothesis

I have argued that when in the delusional condition, the Capgras patient has decoupled the conflict monitor from executive control. This generates the hypothesis that, when in the delusional condition, there will be observable activation of the dACC without the normally expected activation of the PFC. Neuroimaging results should reflect this. In fact the result should be the same, or very similar, to those produced during investigation of subjects in the hypnotic condition. Testing Capgras patients presents unique challenges because the delusion is stimulated by viewing the face of a particular loved-one. Using photographs is not feasible because there has been no recorded delusional response to photographs of the target loved-one. Therefore the loved-one that stimulates the delusion needs to be present during testing. Despite the difficulties, testing should be possible using an open structure MRI scanner. A second prediction generated by my hypothesis is that patients, who are vulnerable to the Capgras delusion, are those that are able to decouple their monitoring system from executive control more readily and completely than others. As this would be reflected in their ability to become more absorbed in alternative realities than others, patients should score higher on the Multiple Personality Questionnaire Absorption Scale.

> Persons who score high on absorption are described as those who are emotionally responsive to engaging sights and sounds, are readily captured by entrancing stimuli, and think in images. Such individuals seemingly can summon and become absorbed in vivid and compelling recollections and imaginings and can experience episodes of expanded awareness. (Menzies, Taylor & Bourguignon: 2008, 300)

In contrast, subjects with low absorption capacity are not easily caught up in sensory or imaginative experiences and "do not readily relinquish a realistic frame of reference" (Menzies, Taylor & Bourguignon: 2008, 300).

CHAPTER 11: FURTHER APPLICATIONS

1. Introduction

I have explicated my hypothesis about delusions using the Capgras delusion as my primary example. In this chapter I will explain how the hypothesis might be applied to five other monothematic delusions. I tentatively suggest three of the delusions have the same putative first and second factor as the Capgras delusions. The first factor is a delay in cognitive processing, with an attribution drawn from a fictional reality. The second factor is the automatic decoupling of a conflict monitoring system from executive control. The three delusions are mirrored-self misidentification without mirror agnosia, thought insertion and thought control. The other two delusions – mirrored-self misidentification with mirror agnosia and somatoparaphrenia —have a different aetiology. In these delusions I propose that the first factor is the developmental impairment of a cognitive skill rather than a processing delay, with the attribution of meaning to the experience involving a non-fictional alternative reality. Nevertheless, they still require the second factor, that is, the decoupling of a conflict monitoring system and executive control, to resolve cross-world confusion, allowing an anomaly to pass unchallenged.

2. The alternative reality of the imagination

i) Mirrored-self misidentification without mirror agnosia

Mirrored-self misidentification is the delusional belief that one's reflection in the mirror is a stranger who may or may not look like oneself. It "almost always occur[s] in patients with focal brain damage or general cerebral dysfunction" (Villarejo *et al.*: 2011, 276). Test results from two patients (FE and TH) led Breen, Caine and Coltheart (2001) to hypothesise that there are two possible causes of mirrored-self misidentification. They suggest one delusion results from a face processing deficit and the other from an inability to interact appropriately with mirrors (mirror agnosia). Mirrored-self misidentification without mirror agnosia is the delusion that purportedly results from a deficit in

face processing. Test results for FE and TH showed that TH suffered from mirror agnosia but FE did not. Significantly, the results for FE "indicated poor structural encoding of faces" (Breen, Caine & Coltheart: 2001, 249). If the 'poor structural encoding' involves a delay in face encoding, it might well create a data conflict (self/not-self) in the same way as I hypothesise the known/unknown conflict is created in the Capgras delusion. A delay in the usually rapid processing of the patient's own face would lead to the false classification of the face as 'not self'. However, as with the Capgras delusion, when an identity match is found for the face (self), it creates a data conflict. Whether the face is judged to be that of a stranger or that of a close relative might be dependent on the speed of the processing. If the processing is sufficiently slowed, it could result in an 'unfamiliar' or 'stranger' classification. A processing speed fast enough for the face to be classified as familiar, but not fast enough for the face to be classified as 'self', might lead to the judgment that the face belongs to a close relative. Van den Stock and colleagues (2012) describe a case of a woman who misidentified herself in mirrors. Test showed she had selectively impaired face-familiarity recognition.

> Although the delusion only pertained to her own face, the face-familiarity deficits also applied to other faces, including close relatives. Possibly, MV relied on non-face cues like body or voice recognition to recognize familiar persons, whereas self-identification in the mirror primarily relies on face-cues. (Van den Stock et al.: 2012, 3)

Their speculation is in keeping with my suggestion in Chapter 6 that the use of cues enables patients with poor face processing skills to encode a face rapidly enough to correctly classify it as familiar, and that the removal of those cues can result in delusion.

As patients understand how mirrors work, yet attribute an alternative identity to their reflection, the attribution needs to be sourced from, or at least linked to, a reality in which this is possible. Seeing something other than one's own reflection when viewing oneself in a mirror is not consistent with everyday reality. It is only in tales of imagination that one's reflection might be replaced by another identity. For example, when Harry Potter looked in the mirror he saw his deceased parents replacing his reflection, and the jealous queen in Snow White saw the face of the talking mirror in one version of the story, and Snow

White's face in another. If the meaning attribution is sourced from fiction, the cross-world confusion it creates requires the same putative involuntary decoupling of the conflict monitor from executive control (to resolve the conflict) that I propose is the case with the Capgras delusion. It gives rise to the same prediction regarding brain activations, that is, that there will be activation of the anterior cingulate cortex without the accompanying activation of the lPFC that is purported to reflect the engagement of executive control subsequent to the activation of a conflict monitor. A second prediction is that patients will score higher than normal subjects on the Multidimensional Personality Questionnaire Absorption Scale.

Using neuroimaging to test the veracity of the first prediction faces challenges. The neural correlates of visual self-recognition differ from those of other face recognition and "there is a lack of convergence as to the precise neuroanatomical locations underlying self-face recognition" (Devue et al.: 2007). For example, despite the many studies that have been conducted, "as far as hemisphere dominance of self-face recognition is concerned, the conclusions of the studies are contradictory" (Devue & Brédart: 2011, 42). Studies are further complicated by the wide range of areas that have been implicated in self-face recognition, and also by a lack of distinction between self-face recognition and self awareness. However, what has been established is that self-face recognition utilizes some areas of the brain that differ from those employed during the recognition of other highly familiar faces (Devue et al.: 2007). "Involvement of different parts of the cingulate cortex have been consistently reported during self-face recognition by comparison with familiar faces recognition, namely, the anterior part of the right" (Devue & Brédart: 2011, 46). The involvement of the anterior cingulate cortex suggests a possible overlap with the proposed conflict monitoring area. Having said that, further research is needed to clarify the situation, because studies conducted by cognitive neuroscientists in an attempt to identify the neural processes underlying participants responses to images of their own faces have all used photographs rather than mirrors. Recent studies show that "self-recognition in different

media involves distinct neural signatures in relation to the featural, configural, and matching stages of face recognition" (Butler et al.: 2012, e31452). Thus they conclude that "generalisations about the brain processes underlying self-recognition based solely upon studies using photographs may not be warranted" (Butler et al.: 2012, e31452). The need for more research in this area is clearly indicated.

ii) Thought insertion

Thought insertion is the belief that thoughts generated by an external agent are being inserted into one's mind. The alien thoughts are usually attributed to one of three sources: other people, machines, or supernatural beings. Researchers have argued for a wide range of disturbances and influences that might explain the condition. Some examples are: loss of awareness of intention to think (Frith: 1992), loss of sense of agency (Stephens & Graham: 1994; Sousa & Swiney: 2011), loss of self-knowledge (Bortolotti and Broome: 2009; Fernández: 2010), a delay in corollary discharge (Whitford et al.: 2011; Feinberg, I: 2011), a disturbance in cue integration in the experience of agency (Moore & Fletcher: 2012), the influence of metacognitive beliefs (Morrison, Haddock & Tarrier: 1995) and the influence of cultural beliefs on sensory override (Luhrmann: 2011).[71] Research into the aetiology of thought control continues.

My suggestion is that thought insertion also might be caused by a processing delay. However, in the case of thought insertion, the processing delay is not in face encoding but in the production of inner speech.

> Most people hear a little voice inside their head when thinking, reading, writing, and remembering. This voice is inner or internal speech, mental imagery that is generated by the speech production system. (Oppenheim & Dell: 2008, 528)

Translating the thoughts of others into inner speech when reading normally involves visual processing, and therefore, such thoughts are not processed by

[71] Luhrmann notes that an attempt to translate the concept of thought insertion and withdrawal in Borneo, failed. "The Iban do not have an elaborate idea of the mind as a container, and so the idea that someone could experience external thoughts as placed within the mind or removed from it was simply not available to them" (Luhrmann: 2011, 70).

our minds as fluently as those we produce ourselves. Understanding the spoken thoughts of others involves auditory processing, which also takes longer than producing the thoughts ourselves. Consequently, a slight slowing in the production of one's own inner speech could cause it to be classified as belonging to an external agent. But in this case, there is no obvious source of the thoughts – no book, no radio, no television or talking person. The classification of the thoughts as being externally generated in the absence of any identifiably external source creates an anomaly. Again, I suggest the anomaly is resolved by sourcing the meaning component of the immediate perceptual experience from an alternative reality.

Receiving thoughts from others (telepathy) and communications from spiritual beings are familiar notions in our culture. In fact, Mullins and Spence note that these ideas have become culturally acceptable.

> Certain phenomena resembling thought insertion have gained cultural credence through being incorporated into occult, parapsychological and religious literature. (Mullins: & Spence: 2003, 293)

The popularity of New Age beliefs and the increase in charismatic religions have made the idea of communicating with spiritual or alien agents, widespread. This is important because "the religious system in which people are embedded will shape what is possible in their experience" (Luhrmann: 2011, 76). The cultural prevalence of beliefs in these forms of externally sourced thoughts means that most people are familiar with the ideas even if they regard them as fiction. Whichever the case, the 'supernatural being' or 'telepathy' attribution is sourced from an alternative reality, either that of the paranormal or that of fiction, and it requires the putative decoupling of the conflict monitoring system from executive control to allow the experience to pass unchallenged. So long as the particular experience is benign and in keeping with cultural norms, it is no more of a problem than is the decoupling that takes place to allow someone to enjoy reading a novel. [72] The experience only becomes problematic when thoughts are

[72] Subjects who hold strong religious or occult beliefs in keeping with cultural norms are not classified as delusional unless there are other reasons for doing so.

intrusive, malicious, persistent, or disturbing in some other way, and the patient is unable to prevent their occurrence.

Not all meaning attributions designed to account for the lack of an identifiable source of the external thoughts fall under the religious/parapsychology umbrella. Some patients claim the thoughts are produced by inanimate objects, for example, household appliances (Jackson & Fulford: 1997) or an electrical machine (Tausk: 1988). Arguably, the belief that the thoughts are being inserted by some kind of mechanical object is drawn from science fiction. But whether fictional, religious or parapsychological, such meaning attribution involves an alternative reality, and the cross-world conflict it creates should attract the attention of the conflict monitor. The resulting triggering of executive control would then stimulate the re-examination of the data. But, as with the Capgras delusion, I suggest re-examining the data merely generates new aberrant experiences. The problem is ultimately resolved by the involuntary decoupling of the conflict monitor and executive control to allow the meaning to pass unchallenged. Whether or not this constitutes a problem is dependent on one's pre-existing beliefs and the norms of one's culture.

The above speculations again give rise to the prediction that there will be an activation of the brain areas designated as the location of a conflict monitoring system, without the subsequent engagement of executive control. Patients experiencing thought insertion should also score high on the Multidimensional Personality Questionnaire Absorption scale.

iii) Thought control

Thought insertion and thought control are usually classified as two different delusions, but it could be argued that they are both based on the same anomaly with the difference being dependent on what aspect of the thought the patient finds most salient. In one case it may be the content of the thought that is salient, and in the other, the inability to control the content. If, for example, the content of the thought is perceived as not being in keeping with the patient's

self image, she may respond to the conflict between her self-image and the thought by rejecting ownership of the content. Thus, she comes to believe that the external agent is *inserting* the thought into her mind. Alternately, a patient who holds prior beliefs about the importance of being able to control her thoughts (common in a wide range of philosophies and religions), the most salient aspect might be the lack of coherence between this belief and her inability to control her thoughts. Thus, she might come to believe the external agent is *controlling* her mind and forcing her to produce the unwanted thoughts. In both cases, the underlying anomaly could be a slowing of thought processing, which results in the false classification of the patient's inner speech as being externally generated, with the content of the delusional belief reflecting different ways in which the anomaly is interpreted in relationship to prior beliefs.

3. The alternative realities of cognitive levels

Whether or not religious experiences are classified as a separate reality or merely a genre of fiction depends on one's personal point of view. Less controversial are the different realities we experience as children during the various stages of cognitive development. How the world looks depends on the cognitive skills we have developed and the concepts we have acquired. At any stage in our development, brain impairment can result in cognitive regression and the partial return to earlier patterns of reacting or thinking. According to Whittlesea and Williams, a monitoring system detects any incoherence or discrepancy that occurs between cognitive processes. If one aspect of a person's environment is interpreted from a different cognitive level of operation to that used to interpret the rest of her environment, the putative monitoring system should register a lack of coherence. I tentatively suggest that the following two delusions result from cross-world confusion caused when the mind attempts to integrate data from two different cognitive levels of operation. More specifically, the problem occurs when an earlier pattern of thinking that occurs in response to a stimulus does not cohere with the general level of thinking operating at the same time. Attempting to resolve the conflict by examining the environment

more closely merely serves to recreate the anomalous experience, so to resolve the conflict, the conflict monitor automatically de-couples from executive control.

i) Mirrored-self misidentification with mirror agnosia

Breen, Caine and Coltheart (2001) hypothesize that mirrored-self misidentification can result from a loss of the ability to interact appropriately with mirrors. Prior to the age of 18-24 months children do not understand how mirrors work. To do so the child has to learn to negotiate the two spatial worlds involved in the *subjective*, kinaesthetic feeling of her body and the *objective*, visual experience of her reflected image. "Learning to recognise one's own reflected image is, then, a singular cognitive achievement, in both developmental and evolutionary terms" (Caine: 2009). The similarity between patients suffering mirrored-self misidentification with mirror agnosia and young children has been noted.

> Ajuriaguerra et al. (1963) compared their patients' interactions with mirrors to those of infants and suggested that the deterioration in their patients' use of mirrors may loosely follow the reverse sequence of stages that infants go through when they learn how to use mirrors. (Connors & Coltheart: 2012)

According to the child's understanding, the face in the mirror belongs to another person. When adults lose their understanding of how mirrors work the reflected face is again seen as that of another person. Both children and adult patients talk to the person in the mirror, but while adults frequently find the fact that the stranger does not respond puzzling and often disturbing, children react to the image as to that of a sociable playmate. The difference between the child and the delusional adult is that whilst the child's interaction with mirrors is in keeping with the rest of her cognitive operations, the adult's interaction is not, thus creating dissonance between the two levels of operation. If existential tagging identifies the 'stranger' in the mirror as belonging to a way of *being-in-the-world* consistent with a child perspective, rather than the way of *being-in-the-world* the adult generally experiences, the conflict monitor should detect a discrepancy and alert executive control. However, as with the Capgras delusion, re-examining the data only served to generate new anomalous experiences. To

resolve the problem, I suggest the conflict monitor involuntarily decouples from executive control to enable the 'stranger in the mirror' experience to pass unchallenged. As with the other delusions I have described, the above gives rise to the prediction that a lack of engagement of executive control after the activation of the conflict monitor would be reflected in neuroimaging data.

ii) Somatoparaphrenia

Patients suffering somatoparaphrenia claim that a body part, usually their arm or hand, belongs to another person. A review of the neuropsychological literature (Vallar & Ronchi: 2009) shows that the other person is predominantly the doctor or the examiner, but nurses, relative and friends are also identified as the owner of the body part.[73] The delusional misidentification of body parts (somatoparaphrenia) is secondary to asomatognosia (unawareness of ownership of body parts). It is typically associated with unilateral neglect (a deficit in attention to and awareness of one side of space), and most frequently with anosognosia for hemiplegia (denial of paralysis). "This association of symptoms makes anatomical inference based on single case studies not sufficiently specific" (Gandola et al.: 2012, 1165). There is a need to clarify which neural networks are involved with each of the specific conditions. In an attempt to do so, Todd Feinberg, and colleagues (2010) studied three groups of patients with right hemisphere strokes and left hemiplegia. The three groups were those with asomatognosia and neglect, non-asomatognosia and neglect, and hemiplegia only. The asomatognosia group was further subdivided into somatoparaphrenia (asomatognosia with delusion) and simple asomatognosia (asomatognosia without delusions). They concluded from their studies that asomatognosia results from large lesions involving multiple sectors of the brain, with medial frontal involvement appearing to be important. Somatoparaphrenia is distinguished from simple asomatognosia by the inclusion of an additional lesion in the orbitofrontal area. In further research, Invernizzi et al. (2012)

[73] It could be argued that the predominance of doctors and examiners nominated as owners merely reflects the fact that they are immediately present when the question (e.g. whose hand is this?) is asked.

described five patients with somatoparaphrenia and mild unilateral neglect, but without anosognosia for hemiplegia. They noted that the lesion distribution of each patient clearly shows that the pure somatoparaphrenia patients had a sparing of most of the regions associated with anosognosia for hemiplegia. They suggest that this new anatomical evidence indicates that motor awareness is not sufficient to build up a sense of ownership.

Theories of the cause of somatoparaphrenia have concentrated on the loss of awareness of ownership (asomatognosia) that is necessary before ownership of the body part can be attributed to someone else. A widely accepted theory is that the loss of ownership results from a disturbance in the multisensory integration thought to be necessary for the experience of body ownership.

> The perception of one's hand is an eminently coherent impression: proprioceptive, tactile, and visual inputs usually correspond perfectly. Current theoretical frameworks postulate that this multimodal integration is required for a feeling of ownership of the body, and is ultimately accompanied by a sense of self. (Bekrate-Bodmann, Foell, & Kamping: 2011, 9443)

The relationship between multisensory integration and body ownership is clearly demonstrated by the 'rubber hand' experiment. The experiment was first conducted by Botvinick and Cohen in 1998. They describe their experiment as follows.

> Each of ten subjects was seated with their left arm resting upon a small table. A standing screen was positioned beside the arm to hide it from the subject's view and a life-sized rubber model of a left hand and arm was placed on the table directly in front of the subject. The subject sat with eyes fixed on the artificial hand while we used two small paintbrushes to stoke the rubber hand and the subject's hidden hand, synchronising the timing of the brushing as closely as possible. (Botvinick & Cohen: 1998, 756)

Subjects reported that during the experiment they experienced the illusion of touch on the rubber hand and also experienced the rubber hand as belonging to themselves. A wide range of 'rubber hand' experiments have been devised and performed since the publication of this experiment. More recently Lopez and colleagues (2012) successfully conducted a non-visual variant of the rubber hand illusion by applying tactile stimuli with temporal-spatial congruency. During the experiment subjects experienced ownership of the rubber hand and felt they were touching their left hand with their right index finger. The above experiments support the idea that the loss of a sense of limb ownership that

underlies somatoparaphrenia could be caused by a disturbance in multisensory integration.

Researchers have made two intriguing discoveries regarding somatoparaphrenia. Somatoparaphrenia can be momentarily reversed both by vestibular stimulation, and by the patient viewing herself in a mirror. It is possible that the first could result from the influence of vestibular stimulation on visual capture. That is, it "might be due to a spatial or temporal modification of visual-tactile integration, leading to an enhancement of visual capture" (Lopez, Lenggenhager & Blanke: 2010, 45), and as a result, the dominance of visual data over tactile data. The second, the fact that mirror-view momentarily reverses somatoparaphrenia, led Fotopoulou and colleagues (2011) to suggest there is a dissociation of the neurocognitive mechanisms that support first-person and third-person visual perspectives on body ownership.[74] They speculate that somatoparaphrenia is caused by two impairments. The first is impairment to the first-person visual perspective.

> The contribution of the first-person perspective to self-identification seems to be impaired in somatoparaphrenic patients, since they fail to identify their contralesional limbs as their own when they observe them directly. (Fotopoulou et al.: 2011, 3947)

They suggest this impairment results in a dominance of the 'felt' body over the 'seen' body.[75] Repeated alternation between the third-person (mirror) perspective, in which ownership was exhibited, and the first-person perspective, in which it was not, failed to elicit a permanent remission in the condition.

> This suggests impairment in the integration and higher order re-representation of various bodily signals into a coherent representation of body ownership, involving both embodied and disembodied perspectives. We propose that somatoparaphrenia can be regarded as a neurogenic dissociation between the felt and the seen body, reminiscent of how infants are not fully able to integrate their 'felt body' with their disembodied image in the mirror. (Fotopoulou et al.: 2011, 3953)

[74] They note that mirrored-self misidentification also indicates this type of dissociation.

[75] If Fotopoulou and colleagues are correct it might explain why vestibular stimulation, which enhances visual capture, briefly reverses the condition. Arguably, it would increase the strength of the 'seen' body, thus working against the dominance of the 'felt' body.

Dysfunction of integration and higher order re-representation constitutes the second impairment.

Whereas infants fail to integrate their 'felt' body with their 'seen' body in the mirror (third-person perspective), patients suffering somatoparaphrenia appear unable to integrate their 'felt' body with the 'seen' body at the first-person perspective. It might be speculated that the apparent dominance of the 'felt' body over the 'seen' body in somatoparaphrenia reiterates the prenatal experience, in which the tactile dominates the visual in foetal intrauterine behaviour (Mori: 2010). Impairment of the first-person visual perspective might force patients to fall back on earlier experiential perspectives. However, as Fotopoulou and colleagues note, a further factor is necessary to explain the persistence of the delusion in the face of evidence to the contrary. This makes their account a three-factor account.

The similarities between the characteristics of somatoparaphrenia and early childhood perception suggest that the condition might involve a loss of the ability to utilise concepts acquired during cognitive development. To believe, for example, that one's arm belongs to someone else requires the loss of the concepts of both body form and body integrity. The concept of body form is the awareness that the human body has a particular form (one head, two arms, and so forth) coupled with an awareness of how the parts relate to one another. The concept of body integrity is the awareness that despite its ability to grow and change, the body itself is an enduring and continuous whole. There comes a time when an adult pretends to steal the child's nose and she no longer needs to check to see if it is still there (Feiner: 1987). The development of anatomical constancy occurs at about the same time as the formation of invariance occurs in other cognitive realms, that is, between the ages of five and seven (Feiner: 1987). Normally, when there is a loss of sensory feedback from any part of the body, the attribution of meaning to the experience is constrained by the concepts of body form and integrity. For example, when we have a dental injection we do not believe that part of our face has disappeared, despite how it feels. That the patient demonstrates a loss of the constraining concepts of body

form and integrity with regard to the disowned body part suggests that she perceives the experience from an earlier stage of cognitive development. That is, she perceives it through the eyes of a child, and the meaning content of the immediate perceptual experience will be commensurate with that view.[76]

What I tentatively suggest is that, as a result of one or more of the widespread lesions that have been identified in somatoparaphrenia, the patient is unable to access the concepts of body form and integrity. As with those suffering mirrored-self misidentification, described above, the patient comes to view one aspect of her world from a different conceptual perspective to that of the rest of her world. More specifically, the patient interacts with her estranged limb from a conceptual level in keeping with that of a child, whilst interacting with the rest of her world as an adult. If we take the different cognitive developmental levels to be different realities, the inserting of a child perspective into the normal adult perspective creates cross-world confusion and should attract the attention of the conflict monitor system. Again, as I have hypothesised with the Capgras and other delusions, checking the environment only serves to renew the anomalous experience, and the conflict is ultimately resolved by a second factor, the automatic decoupling of the conflict monitor from executive control. The ease

[76] The constraining influence of body concepts has led to the assumption that the rubber hand experiment cannot be used to create the experience of supernumerary limbs. However, several studies report the duplication of the right hand in subjects exposed to two adjacent hands. Folegatti and colleagues (2012) tested whether or not spatial constraints affect the possible inducing of the sense of ownership in two rubber hands located side by side. They found that only the closest rubber hand appeared both objectively (proprioceptive drift) and subjectively (ownership rating) embodied. To illicit a genuine illusion of multiple hands the rubber hands must be at the same distance from the subjects hand/body. They argue "the brain computes the boundaries of the biological body, but taking into account some margin of spatial errors, thus creating a gray zone of possibilities of body ownership" (Folegatti et al.: 2012, 811). Rubber hand experiments create the reverse ownership anomaly to somatoparaphrenia. But any demonstration that body concept does not restrain the illusion of anomalous limb numbers in healthy subject may be relevant to the delusion. In this case, Folegatti and colleagues' conclusion that ownership of supernumerary limbs occurs in a gray zone of ambiguity is not consistent with delusional patients' attribution of ownership to other people who may be at a considerable distance from the patient's own body (e.g. in another room).

with which a particular patient is able to execute the decoupling would determine the strength and persistence of the delusion.

4. Schizophrenia

Schizophrenia is not the subject of this thesis, and, therefore, the following suggestions I make are very broad, general and highly speculative. Schizophrenia is not a clearly defined concept. The essential features of schizophrenia are a mixture of characteristic signs and symptoms that may be positive or negative[77].

> The characteristics of Schizophrenia involve a range of cognitive and emotional dysfunctions that include perception, inferential thinking, language and communication, behavioral monitoring, affect, fluency and productivity of thought and speech, hedonic capacity, volition and drive and attention. No single symptom is pathognomonic of Schizophrenia; the diagnosis involves the recognition of a constellation of signs and symptoms associated with impaired occupational or social functioning. (DSM-IV-TR: 2005, 299)

As schizophrenia is defined as a 'constellation of signs and symptoms' it is unlikely that there is a single underlying cause. My suggestion that the decoupling of the conflict monitoring system from executive control is involved, might apply to some symptoms but is not likely to apply to others. A likely candidate for my hypothesis is auditory hallucination. Auditory hallucinations are usually experienced as 'voices' that are perceived as distinct from the person's own thoughts, and, according to the DSM, are by far the most common of the sensory hallucinations[78]. Neuroimaging studies suggest that the auditory cortex is more active during the hallucinatory experience than at times when the hallucination is not present.

> This finding could be interpreted as increased spontaneous neuronal activity, or hyper-excitability of the auditory cortex. Spontaneous fluctuations of activation in sensory cortices have been shown to affect stimulus processing. (Kompus, Westernhausen & Hugdahl: 2011, 3361)

The area of the auditory cortex that is spontaneously activated is the area that normally responds to the voice of others, that is, the brain responds 'as if' it is hearing someone speaking. No matter how closely the environment is

[77] Positive symptoms are thought to reflect an excess or distortion of normal functions, whilst negative symptoms reflect a diminution or loss of normal functions.

[78] Auditory hallucinations are not always associated with schizophrenia.

examined, a source for the voice will not be found. A voice that has no identifiable source is not in keeping with everyday reality, and therefore, accounting for it is likely to involve an alternative reality. The cross-world conflict this creates will stimulate the second factor, as previously described. The hallucination will become a delusion to the degree to which the conflict monitor and executive control decouple.

A feature of schizophrenic delusions is the bizarre nature of their content—content that frequently does not comply with the constraints of the real world. According to the DSM, "bizarre delusions are considered to be especially characteristic of Schizophrenia" (DSM-IV-TR: 2005, 299). For Jaspers, the delusions "remain largely incomprehensible, unreal and beyond our understanding" (Jaspers: 1963, 98), and he describes them as belonging to a 'world of new meaning'. However, according to Currie (2000), we would not find the thoughts of those suffering schizophrenic delusions so bizarre if we viewed them as flows of imaginings that the patient mistakes for beliefs. My suggestion is that we would not find them so bizarre if we viewed them as interpretations of the world drawn from an alternative reality. Sass reports that a schizophrenic patient (Schreber) claimed that people would sometimes change heads with each other. According to Sass, "He seems to have experienced such events as occurring in a realm distinct from the natural or actual world, with its constraints and consequentiality" (Sass: 1994, 46). As with other delusions that are part of an elaborated belief web, it is difficult to determine the precise reality from which such an experience might be drawn. The delusion that people were exchanging heads (intermetamorphosis) might have involved the loss of the concepts of body form and integrity (child perspective) or merely reflected a science fiction fantasy. Again, when Schreber claims "everything that happens is in reference to me" (Sass: 1994, 56), it is reminiscent of the egocentricity of early childhood. However, considering Schreber believed he had a special relationship with God (he was the conduit through which God can know), and that he had become the centre of divine miracles, his belief that everything happened in reference to him could have been a religious delusion.

The elaborated, polythematic delusions characteristic of schizophrenia, appear to involve a range of alternative realities, with the content being drawn from a variety of genres of fiction, fantasies, reveries and the like, and childhood or religious perspectives. According to my account of the Capgras delusion, in an attempt to resolve a cross-world conflict, the conflict monitoring system is decoupled from executive control. Two brain areas implicated in these processes are the anterior cingulate cortex (conflict monitoring) and the dorsolateral prefrontal cortex (executive control). Empirical data show that patients with schizophrenia have impairment at both these sites. "Executive control is an aspect of cognitive function known to be impaired in schizophrenia" (Gilmour et al.: 2012, 1). "The dysexecutive syndrome is one of the most prominent and functionally cognitive features of schizophrenia" (Raffard & Bayard: 2012, 60). In addition, "morphological abnormalities of the anterior cingulate occur in patients with schizophrenia and in symptomatic high risk individuals" (Meredith et al.: 2012, 377). Anterior cingulate volume abnormalities appear to be a correlate of increased vulnerability to psychosis (Rothlisberger et al.: 2012). Without addressing the issue of what the first factor might be, or attempting to describe the mechanism, I make the broad suggestion that, rather than accepting one bizarre meaning that results from a specific deficit (as I propose occurs in some monothematic delusions), a general deficit in conflict monitoring and/or executive control could result in all manner of bizarre meanings attached to the immediate perceptual experience to pass unchallenged or uncorrected. Without a strong conflict monitoring system and a well functioning executive control system, the schizophrenic patient might be left drifting between realities without a means of determining which world is real.

SUMMARY AND CONCLUSION

In this thesis I have argued that the most plausible account of monothematic delusions ultimately incorporates valuable insights from both the continental or phenomenological tradition and the Anglo-American or analytic tradition. I argue my case by presenting a new two-factor account of the Capgras delusion in which the characterization of each factor reflects insights from both traditions. In doing so I have attempted to demonstrate that an account that draws on insights from both traditions better explains the characteristics of the delusion than accounts that draw from one tradition only. The thesis was divided into two parts. The first reviewed relevant literature on delusions. The second gave a detailed description of the two new factors in a two-factor account of the Capgras delusion that constitutes my hypothesis.

Part One of the thesis began with a description of the views of three influential phenomenologists – Jaspers, Sass and Gallagher. I noted the following ways in which I have been influenced by their work. From Jaspers I have drawn the conception of delusion as a disturbance in the formation of the meaning component of the immediate perceptual experience, and from Jaspers, Sass and Gallagher, the suggestion that delusion involves an alternative reality. More generally, I accepted the phenomenological view that delusion is a disturbed experience that reflects the subject's way of *being-in-the-world.* Therefore, it is not a belief, although the disturbed experience may be endorsed as a belief. Further, I accepted that it is important to consider the whole person, embedded in her environment, rather than restricting research to the study of isolated cognitive processes.

In the second chapter I began my review of the analytic literature, beginning with Maher's challenge to Jaspers, and his influential 'explanationist' account of delusions. I described his claims that delusions are caused by an underlying biological impairment that gives rise to an anomalous experience, and that the patient's explanation of the experience is the result of normal thinking

processes. I acknowledged the empirical data that resulted in the general acceptance of Maher's first claim. I then described various attempts to prove or disprove his contentious second claim. I noted weaknesses in Maher's account that led researchers to conclude that his account provides insufficient explanation of the formation and maintenance of monothematic delusions, and that a second factor is necessary. In the third chapter I introduced two-factor accounts, distinguishing between two-factor and two-deficit accounts. I described and discussed two two-factor accounts: Coltheart's (2007) two-deficit account, and Davies and Davies (2009) parametric variations account. As some researchers consider monothematic delusions can be plausibly explained by a single factor, the following chapter was devoted to describing and discussing two one-factor accounts – Gerrans' (2009) dopamine account, and Corlett, Fletcher and colleagues' prediction error account.

In the final chapter of Part One, I discussed two contentious issues important to any discussion of delusions. The first was the doxastic status of delusions and the degree to which experience determines the content of the belief. The second issue concerned the poverty or richness of perceptual content. I began discussing the first by acknowledging three non-doxastic accounts – those of Berrios (1991), Currie (2000), and Sass (1992) – but noted that, in keeping with the DSM definition of delusions, most analytic researchers regard delusions as beliefs. I then described two ways in which beliefs are thought to be formed. The first was the explanationist account, which supports Maher's contention that the belief is an explanation of an anomalous experience. The second was the endorsement account, which holds that the belief is formed by accepting an anomalous experience as veridical, and is, therefore, more aligned to the phenomenological view. I described two more recent accounts – those of Coltheart, Menzies and Suttons (2010) and Langdon and Baynes (2010) – which attempt to harness the strengths of the explanationist and endorsement accounts, whilst avoiding their weaknesses. I acknowledged that my account is an endorsement account.

In the second section of the final chapter in Part One I discussed the issues that arise when attempting to ascertain the contents of perception. I described three competing accounts of perceptual content – those of Fodor (1983), Churchland (1988) and Raftopoulos (2001). I then described Gallagher's more liberal interpretation of the term 'perception', which yields a far richer perceptual content than those previously discussed. In the account of the Capgras delusion I offered, I argued for an extremely rich perceptual content. In doing so I drew on Gallagher's conception of perception as being direct, and the product of an embodied, embedded and enactive cognition.

Part Two of the thesis began with an account – the delayed-processing account – of a new first factor in a two-factor account of the Capgras delusion. I hypothesised that a slowing of face encoding causes the face of a loved-one to be classified erroneously as unfamiliar, and that a conflict then arises when the face is identified as that of a particular person. I argued that the conflict is resolved by the attribution of the 'impostor' concept, and that the rich content is an interpretation of the face in context. I argued further, that the impostor concept provides the gist of the meaning of the scenario, and constitutes the meaning component of the immediate perceptual experience.

Drawing from Gallagher's alternative realities hypothesis, I then argued that the 'impostor' concept that provides the content of the Capgras delusion is, for most people, formed from engagement in fictional narratives. Therefore, the neural assemblies connecting the concept to a fictional reality will be dominant. Thus the rapid, non-reflective processes that select concepts to provide meaning in the immediate perceptual experience will automatically select the 'impostor' concept linked to the web of beliefs associated with a fictional reality. In further support of the hypothesis that delusions involve fictional realities, I argued that such involvement helps to explain some puzzling characteristics of the Capgras delusion.

The application of a concept tagged as fictional to a person who exists in the real world creates a cross-world conflict. I described the strategy for resolving this conflict as the second factor in the Capgras delusion, and gave a coherence-monitoring account of the factor. Based on empirical data from hypnosis research, I argued that the normal mechanism that enables a subject to enter an alternative reality is the decoupling of the conflict monitoring system and executive control. I applied this to the Capgras delusion by suggesting that the monitoring system and executive control automatically and involuntarily decouple to allow the cross-world attribution to pass unchallenged. According to my hypothesis, the first and second factors together produce the automatic, non-reflective interpretation of the gist of the scenario that provides the meaning content of the immediate perceptual experience.

In the final chapter of Part Two I explain briefly how my hypothesis might be applied to five other monothematic delusions. I suggest that three of these delusions involve an alternative fictional reality, whilst two involve the alternative reality of earlier levels of cognitive development. I also make the broad suggestion that neuropsychological impairment in patients suffering schizophrenia may cause them to have general difficulty in distinguishing alternative realities from the everyday world. I nominated two possible sites of impairment that could result in such confusion.

BIBLIOGRAPHY

ABC Lateline: 2010 *Jonathan Franzen discusses his new book* Leigh Sales (presenter). Lateline Tuesday November 30th. www.abc.net.au/lateline

Ajuriaguerra, J., Strejilevitch, M., & Tissor, R.: 1963 A propos de quelques conduites devant le miroir de sujets atteints de syndromes démentiels du grand âge (On the behaviour of senile dementia patients vis-à-vis the mirror). *Neuropsychologia, 1*, 59-73.

Alexander, M. P., Stuss, D. T., & Benson, D. F.: 1979 Capgras Syndrome: A Reduplicative Phenomenon. *Neurology, 29*, 334-339.

Amiola, A. M.: 1999 Dark Side of the Moon: Studies in Unilateral Neglect. PhD Dissertion, University of Auckland

Anderson, C. A., Camp, J., & Filley, C.: 1998 Erotomania After Aneurysmal Subarachnoid Hemorrhage: Case Report and Literature Review. *Journal of Neuropsychiatry, 10*(3), 330-337.

Anderson, D. N.: 1988 The Delusion of Inanimate Doubles. *British Journal of Psychiatry,* 153, 694-699.

Andreasen, N.: 1988 Brain Imaging: Applications in Psychiatry. *Science, New Series, 239*(4846), 1381-1388.

Antérion, C., Convers, P., Desmales, S., Borg, C., & Laurent, B.: 2008 An odd manifestation of the Capgras syndrome: Loss of familiarity even with the sexual partner. *Clinical Neurophysiology, 38*, 177-182.

Attewell, J., Cox, R., Barnier, A. J., & Langdon, R.: 2012 Modeling Erotomania in the Laboratory With Hypnosis. *International Journal of Clinical and Experimental Hypnosis, 60*(1 (special issue)), 1-30.

Averbeck, B., Evans, S., Chouhan, V., Bristow, E., & Shergill, S.: 2011 Probabilistic learning and inference in schizophrenia. *Schizophrenia Research, 127*, 115-122.

Barnier, A. J., Cox, R. E., O'Connor, M., Langdon, R., Breen, N., & Turner, M.: 2008 Developing hypnotic analogues of clinical delusions: Mirrored-self misidentification. *Cognitive Neuropsychiatry, 13*, 406-430.

Barnier, A., Dienes, Z., & Mitchell, C.: 2008 How hypnosis happens: new cognitive theories of hypnotic responding. In M. R. Nash and A. J. Barnier (eds.), *The Oxford handbook of hypnosis: Theory, research and practice*. Oxford: Oxford University Press.

Bauer, R.: 1984 Autonomic Recognition of Names and Faces in Prosopagnosia: A neuropsychological Application of The Guilty Knowledge Test. *Neuropsychology,* 22(4), 457-469.

Bayne, T.: 2009 Perception and the Reach of Phenomenal Content. *The Philosophical Quarterly, 59*(236), 385-404

Bayne, T., & Pacherie, E.: 2004 Experience, Belief, and the Interpretive Fold. *Philosophy, Psychiatry & Psychology, 11*(1), 81-86.

Bayne, T., & Pacherie, E.: 2005 In Defence of the Doxastic Conception of Delusions. *Mind & Language, 20*(2), 163-188.

Bekrater-Bodmann, R., Foell, J., & Kamping, S.: 2011 The Importance of Ventral Premotor Cortex for Body Ownership Processing. *The Journal of Neuroscience, 3*(26), 9443-9444.

Bell, V., Halligan, P. W., & Ellis, H. D.: 2008 Are Anomalous Perceptual Experiences Necessary For Delusions? *Journal of Nervous & Mental Diseases, 196*, 3-8.

Bentall, R.: 2003 Madness Explained: Psychosis and Human Nature. London: Penguin.

Berrios, G. E.: 1991 Delusions as 'wrong beliefs': a conceptual history. *British Journal of Psychiatry, 159* (suppl. 14), 6-13.

Blakemore, S.-J., Oakley, D., & Frith, C.: 2003 Delusions of alien control in the normal brain. *Neuropsychologia, 41*, 1058-1067.

Blankenburg, W.: 1969 (Aaron Mishara trans. 2001) First Steps Towards A Psychopathology Of "Common Sense". *Philosophy, Psychiatry & Psychology, 8*(4), 303-315.

Bloom, F., Nelson, C., & Lazerson, A.: 2001 *Brain, Mind, and Behavior* (3rd ed.). USA: Worth Publishers.

Blount, G.: 1986 Dangerousness of patients with Capgras syndrome. *Nebraska Medical Journal, 71*.

Bourget, D., & Whitehurst, L.: 2004 Capgras Syndrome: A Review of the Neurophysiological Correlates and Presenting Clinical Features in Cases Involving Physical Violence. *Canadian Journal of Psychiatry, 49*(11), 719-725.

Bortolotti, L.: 2005 Delusions and the Background of Rationality. *Mind & Language, 20*(2), 189-208.

Bortolotti, L.: 2010 *Delusions and other irrational beliefs*. Oxford: Oxford University Press.

Bortolotti, L., & Broome, M. R.: 2009 A role for ownership and authorship in the analysis of thought insertion. *Phenomenology and the Cognitive Sciences, 8*, 205-224.

Botvinick, M., & Cohen, J.: 1998 Rubber hands 'feel' touch that eyes see. *Nature, 391*, 756.

Botvinick, M., Cohen, J., & Carter, C.: 2004 Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Science, 8*(12), 539-546.

Bowers, K. S.: 1990 Unconscious influences and hypnosis. In J. L. Singer (Ed.), *Repression and Dissociation: Implication for Personality Theory, Psychopathology, and Health* (pp. 143-179). Chicago: Chicago University Press.

Box, G., & Draper, N.: 1987 *Empirical Model Building and Response Surfaces*. New York: Wiley.

Braffman, W., & Kirsch, I.: 1999 Imaginative Suggestibility and Hypnotizability: An Empirical Analysis. *Journal of Personality and Social Psychology, 77*(3), 578-587.

Brakoulias, V., Langdon, R., Sloss, G., Coltheart, M., Meares, R., & Harris, A.: 2008 study involving cognitive behavioural therapy. *Cognitive Neuropsychiatry, 13*(2), 148-165 Delusions and reasoning: A.

Breen, N., Caine, D., & Coltheart, M.: 2000 Models of Face Recognition and Delusional Misidentification: A Critical Review. *Cognitive Neuropsychology,* 17(1/2/3), 55-71.

Breen, N., Caine, D., & Coltheart, M.: 2001 Mirrored-self Misidentification: Two Cases of Focal Onset Dementia. *Neurocase, 7*, 230-254.

Breen, N., Caine, D., & Coltheart, M.: 2002 The role of affect in reasoning in a patient with a delusion of misidentification. *Cognitive Neuropsychiatry, 7*(2), 113-137.

Brigetti, G., Bonifacci, P., Borlimi, R., & Ottaviani, C.: 2007. "Far from the heart far from the eye": Evidence from the Capgras delusion. *Cognitive Neuropsychiatry, 12*(3), 189-197.

Brown, A., Cohen, P., Greenwald, S., & Susser, E.: 2000 Nonaffective Psychosis After Prenatal Exposure to Rubella. *American Journal of Psychiatry, 157*, 438-443.

Bruner, J.: 1990 *Acts of Meaning* Cambridge, MA: Harvard University Press.

Burgess, P., & Shallice, T.: 1996 Confabulation and the Control of Recollection. *Memory, 4*(4), 359-411.

Bürgy, M.: 2008 The Concept of Psychosis: Historical and Phenomenological Aspects. *Schizophrenia Bulletin, 34*(6), 1200 -1210.

Butler, D., Mattingley, J., Cunnington, R., & Suddendorf, T.: 2012 Mirror, Mirror on the Wall, How Does My Brain Recognize My Image at All? *PLoS ONE, 7*(2), e31452.

Caine, D.: 2009 Reflecting on Mirror Self-misrecognition. *Neuropsycholanalysis, 11*(2), 211-226.

Capgras, J., & Reboul-Lachaux, J.: 1923 L'illusion des 'sosies' dans un délire systématizé chronique. *Bulletin de la Société Clinique de Médecine Mentale,* 11,6-16

Cascella, N., Gerner, G., Fieldstone, S., Sawa, A., & Schrellen, D.: 2011 The insula-claustrum region and delusions in schizophrenia. *Schizophrenia Research, 133*(1-3), 77-81.

Chase, V. M., Hertwig, R., & Gigerenzer, G.: 1998 Visions of rationality. *Trends in Cognitive Science, 2*(6), 206-214.

Christodoulou, G. N.: 1977 The Syndrome of Capgras. *British Journal of Psychiatry,* 130, 556-564

Christodoulou, G. N.: 1986 Role of Depersonalization-Derealization Phenomena in the Delusional Misidentification Syndromes. *Bibliotheca Psychiatrica,* 164, 99-104

Churchland, P.: 1988 Perceptual Plasticity and Theoretical Neutrality: A Reply to Jerry Fodor. *Philosophy of Science, 55*, 167-187.

Churchland, P.: 2006 Into the brain: where philosophy should go from here. *Topoi, 25*, 29-32.

Clark, A.: 2012 Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *Behavioral and Brain Sciences, (In Press)*.

Coltheart, M.: 2005 Conscious Experience and Delusional Belief. *Philosophy, Psychiatry & Psychology, 12* (2), 153-157.

Coltheart, M.: 2007 Cognitive neuropsychiatry and delusional belief. *The Quarterly Journal of Experimental Psychology, 60*(8), 1041-1062.

Coltheart, M.: 2010 The neuropsychology of delusions. *Annals of the New York Academy of Sciences, 1191*, Issue: The Year in Cognitive Neuroscience 16-26.

Coltheart, M., Langdon, R., & Breen, N.: 1997 Misidentification syndrome and cognitive neuropsychiatry. *Trends in Cognitive Science, 1*(5), 157-158.

Coltheart, M., Langdon, R., & McKay, R.: 2007 Schizophrenia and Monothematic Delusions. *Schizophrenia Bulletin, 33*(3), 642-647.

Coltheart, M., Langdon, R., & McKay, R.: 2011 Delusional Belief. *Annual Review of Psychology, 62*, 271-279.

Coltheart, M., Menzies, P., & Sutton, J.: 2010 Abductive inference and delusional belief. *Cognitive Neuropsychiatry, 15*(1/2/3), 261-287.

Connors, M., & Coltheart, M.: 2012 On the behaviour of senile dementia patients vis-a-vis the mirror: Ajuriaguerra, Strejilevitch and Tissot (1963). *Cognitive Neuropsychiatry, 17*(2), 151-176.

Conway, M.: 1996 Autobiographical knowledge and autobiographical memories. In D. C. Rubin (Ed.) *Remembering our past: Studies in autobiographical memory.* New York: Cambridge University Press.

Corlett, P. R., Krystal, J. H., Taylor, R., & Fletcher, P. C.: 2009 Why do delusions persist? *Frontiers in Human Neuroscience, 3*(Article 12), 1-9.

Corlett, P. R., Taylor, R., Wang, X.-J., Fletcher, P. C., & Krystal, J. H.: 2010 Towards a Neurobiology of delusions. *Progress in Neurobiology, 92*, 345-369.

Corlett, P. R., Honey, G. D., Krystal, J. H., & Fletcher, P. C.: 2011 Glutamatergic Model of Psychosis: Prediction Error, Learning, and Inference. *Neuropsychopharmacology, 36*, 294-315.

Cox, R. E., & Barnier, A. J.: 2009a Hypnotic illusions and clinical delusions: A hypnotic paradigm for investigating delusions of misidentification. *International Journal of Clinical and Experimental Hypnosis, 57*, 1-32.

Cox, R. E., & Barnier, A. J.: 2009b Selective information processing in hypnotic identity delusion: The impact of time of encoding and retrieval *Contemporary Hypnosis, 26*, 65-79.

Cox, R. E., & Barnier, A. J.: 2010 Hypnotic illusions and clinical delusions: Hypnosis as a research method. *Cognitive Neuropsychiatry, 15*(1/2/3), 202-232.

Currie, G.: 2000 Imagination, Delusion and Hallucination. *Mind & Language, 15*(1), 168-183.

Currie, G., & Jureidini, J.: 2001 Delusion, Rationality, Empathy: Commentary on Davies et al. *Philosophy, Psychiatry & Psychology, 8*(2-3), 159-162.

Davelaar, E.: 2008 A computational study of conflict-monitoring at two levels of processing: Reaction time distributional analysis and hemodynamic responses. *Brain Research, 1202*, 109-116.

Davenport, J. L., & Potter, M. C.: 2004 Scene consistency in object and background perception. *Psychological Science, 15*, 559-564.

Davidson, D.: 1985 Incoherence and Irrationality. *Dialectica, 39*(4), 345 - 354.

Davidson, D.: 1986a A coherence theory of truth and knowledge. In E. Lepore (Ed.), *Truth and Interpretation: Perspectives on the philosophy of Donald Davidson* Oxford: Blackwell.

Davidson, D.: 1986b Deception and Division. In J. Elster (Ed.), *The Multiple Self*. Cambridge: Cambridge University Press.

Davidson, D.: 1998 Who is fooled? In J. P. Dupuy (Ed.), *Self-Deception and Paradoxes of rationality.* Stanford CA: CSLI Publications.

Davies, M.: 2000 Interaction without Reduction: The Relationship between Personal and Sub-personal Levels of Description. *Mind & Society, 1*, 87-105.

Davies, M.: 2009 Delusion and Motivationally Biased Belief: Self-Deception in the Two-Factor Framework. In T. Bayne & J. Fernandez (Eds.), *Delusion and Self-*

Deception. Affective and Motivational Influences on Belief Formation. New York: Psychology Press.

Davies, M., & Coltheart, M.: 2000 Introduction: Pathologies of Belief. *Mind & Language, 15*(1), 1-46.

Davies, M., Coltheart, M., Langdon, R., & Breen, N.: 2001 Monothematic Delusions: Towards a Two-Factor Account. *Philosophy, Psychiatry & Psychology, 8*(2-3), 133-158.

Davies, A., & Davies, M.: 2009 Explaining pathologies of belief. In M. R. Broome & L. Bortolotti (Eds.), *psychiatry as cognitive neuroscience*. Oxford: Oxford University Press.

Davies, A., Davies, M., Ogden, J., Smithson, M., & White, R.: 2009 Cognitive and Motivational Factors in Anosognosia. In T. Bayne & J. Fernández (Eds.), *Delusions and Self-Deception: Affective and Motivational Influences on Belief Formation*. New York: Psychology Press.

Davison, K.: 1989 Neuropsychiatry of localized brain disorders and encephalopathies. *Current Opinion in Psychiatry, 2*, 76-83.

de Gelder, B., Meeren, H. K., Righart, R., Van den Stock, J., van de Riet, W. A., & Tammietto, M.: 2006 Beyond the face, exploring rapid influences of context on face processing. *Progress in Brain Research, 155*, 37-48.

De Neys, N.: 2006 Dual processing in reasoning: two systems but one reasoner. *Psychological Science, 17*, 428-433.

de Pinedo-Garcia, M.: 2008 Beyond persons: extending the personal/subpersonal distinction to non-rational animals and artificial agents. *Biological Philosophy, 23*, 87-100.

De Risio, S., De Rossi, G., Sarcgiapone, M., Carmardese, G., Carli, V., Cuomo, C., et al.: 2004 A case of Cotard syndrome.^{123}I-IBZM SPECT imaging of striatal D_2 receptor binding. *Psychiatry Research: Neuroimaging, 130*, 109-112.

Descarries, L., Bérubé-Carrière, N., Riad, M., Dal Bo, G., Mendez, J., & Trudeau, L.-E.: 2008 Glutamate in dopamine neurons: Synaptic versus diffuse transmission. *Brain Research Reviews*, 290-302.

Devue, C., Collette, F., Balteau, E., Degueldre, C., Luxen, A., Maquet, P., et al.: 2007 Here I am: The cortical correlates of visual self-recognition. *Brain Research 1143*, 169-182.

Devue, C., & Brédart, S.: 2011 The neural correlates of visual self-recognition. *Consciousness and Cognition, 20*, 40-51.

Dietl, T., Brunner, H., & Friess, E.: 2003 Capgras syndrome - out of sight, out of mind? *Acta Psychiatrica Scandinavica,* 108, 460-463.

DiLalla, L. F., & Watson, M. W.: 1988 Differentiation of fantasy and reality: preschoolers' reactions to interruptions in their play. *Developmental Psychology, 24*, 286-291.

Dodhia, R., & Metcalfe, J.: 1999 False Memories and Source Monitoring. *Cognitive Neuropsychology, 16*(3/4/5), 489-508.

Doran, J.: 1990 The Capgras Syndrome: Neurological/Neuropsychological Perspectives. *Neuropsychology, 4*(1), 29-42.

DSM-IV-TR: 2005 *Diagnostic and Statistical Manual of Mental Disorders* 4th Ed Text Revision Arlington: American Psychiatric Association

Duggal, H. S.: 2004 Interictal Psychosis Presenting with Fregoli Syndrome. *Journal of Neuropsychiatry and Clinical Neuroscience, 16*(4), 543-544.

Dulai, R., & Kelly, S.: 2009 A case of the body snatchers. *Current Psychiatry, 8*(8), 56-65.

Edelstyn, N. M., & Oyebode, F.: 1999 A Review of the Phenomenology and Cognitive Neuropsychological Origins of the Capgras Syndrome. *International Journal of Geriatric Psychiatry,* 14, 48-59.

Egner, T., Jamieson, G., A., & Gruzelier, J.: 2005 Hypnosis decouples cognitive control from conflict monitoring processes of the frontal lobe. *Neuroimage, 27*, 969-978.

Egner, T., & Raz, A.: 2008 Cognitive control processes and hypnosis. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford Handbook of Hypnosis: theory, research and practice*. Oxford: Oxford University Press.

Ellis, H.: 2007 Delusions: A suitable case for imaging? *International Journal of Psychophysiology, 63*, 146-151.

Ellis, H. D., & Lewis, M. B.: 2001 Capgras delusion: a window on face recognition. *Trends in Cognitive Science, 5*(4), 149-156.

Ellis, H., Lewis, M. B., Moselhy, H., & Young, A.: 2000 Automatic without autonomic responses to familiar faces: Differential components of covert face recognition in a case of Capgras delusion. *Cognitive Neuropsychiatry,* 5(4), 255-269

Ellis, H., Quayle, A., de Pauw, K., Szulecka, T. K., Young, A., & Kolkiewicz, L.: 1996 Delusional Misidentification of Inanimate Objects: A Literature Review and Neuropsychological Analysis of Cognitive Deficits in Two Cases. *Cognitive Neuropsychiatry,* 1(1), 27-40.

Ellis, H., Young, A., Quayle, A., & De Pauw, K.: 1997 Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society of London Biological Sciences,* 264(1384), 1085-1092.

Ellis, H., & Young, A.: 1990 Accounting for Delusional Misidentification. *British Journal of Psychiatry,* 157, 239-248.

Fei-Fei, L., Lyer, A., Koch, C., & Perona, P.: 2007 What do we perceive in a glance of a real-world scene? *Journal of Vision, 7 (1)* (10), 1-29.

Feinberg, T. E., Venneri, A., Simone, A., Fan, Y., & Northoff, G.: 2010 The neuroanatomy of asomatognosia and somatoparaphrenia. *Journal of Neurology, Neurosurgery & Psychiatry, 81*, 276-281.

Feinberg, I.: 2011 Corollary Discharge, Hallucinations, and Dreaming. *Schizophrenia Bulletin, 37*(1), 1-3.

Feiner, K.: 1987 Development of the Concept of Anatomical Constancy: Part 1. *Psychoanalytic Psychology, 4*(4), 343-354.

Fernández, J.: 2010 Thought Insertion and Self-Knowledge. *Mind & Language, 25*(1), 66-88.

Fine, C., Craig, F., & Gold, I.: 2005a The Explanation Approach To Delusions. *Philosophy, Psychiatry & Psychology,* 12(2), 159-163.

Fine, C., Craige, J., & Gold, I.: 2005b Damned If You Do; Damned If You Don't. The Impasse in Cognitive Accounts of the Capgras Delusion. *Philosophy, Psychiatry & Psychology,* 12(2), 143-151.

Fine, C., Gardner, M., Craigie, J., & Gold, I.: 2007 Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cognitive Neuropsychiatry, 12*(1), 46-77.

Fleminger, S., & Burns, A.: 1993 The Delusional Misidentification Syndrome in Patients with and without Evidence of Organic Cerebral Disorder: A Structured Review of Case Reports. *Biological Psychiatry, 33*, 22-32.

Fletcher, P., & Frith, C.: 2009 Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience, 10*, 48-58.

Fodor, J. A.: 1983 *The Modularity of Mind*. Cambridge MA: MIT Press.

Fodor, J. A.: 1988 A Reply to Churchland's "Perceptual Plasticity and Theoretical Neutrality". *Philosophy of Science, 55*, 188-198.

Fodor, J. A.: 2006 Précis of *The Modularity of Mind*. In J. L. Bermudez (Ed.), *Philosophy of Psychology: Contemporary Readings*. New York: Routledge.

Folegatti, A., Farnè, A., Salemme, R., & de Vignemont, F.: 2012 The Rubber Hand Illusion: Two's a company, but three's a crowd. *Consciousness and Cognition, 21*, 799-812.

Forsti, H., Almeida, O., & Iacoponi, E.: 1991 Capgras Delusions in the Elderly: The Evidence for a Possible Organic Origin. *International Journal of Geriatric Psychiatry, 6*(12), 845-852.

Fotopoulou, A., Jenkinson, P., Tsakiris, M., Haggard, P., Rudd, A., & Kopelman, M.: 2011 Mirror-view reverses somatoparaphrenia: Dissociation between first-and-third-person perspective on body ownership. *Neuropsychologia, 49*, 3946-3955.

Frankish, K.: 2009 Delusions: A two-level framework. In M. R. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience. Philosophical perspectives*. Oxford: Oxford University Press.

Friston, K.: 2011 The history of the future of the Bayesian brain. *Neuroimage, 62*, 1230-1233.

Frith, C.: 1992 *The Cognitive Neuropsychology of Schizophrenia* Hove: Lawrence Eribaum

Frith, C.: 2004 The self in action: Lessons from delusions of control. *Consciousness and Cognition, 14,* 752-770.

Frith, C: 2012 Explaining delusions of control**:** The comparator model 20 years on. *Consciousness & Cognition, 21*(1), 52-54

Frith, C., Blakemore, S.-J., & Wolpert, D.: 2000 Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. B, 355*, 1771-1788.

Frith, C., & Done, D.: 1988 Towards a neuropsychology of schizophrenia. *British Journal of Psychiatry, 153*, 437-443.

Fujii, D., & Ahmed, I.: 2002 Characteristics of Psychotic Disorder Due to Traumatic Brain Injury: An Analysis of Case Studies in the Literature. *Journal of Neuropsychiatry and Clinical Neuroscience, 14*(2), 130-140.

Gallagher, S.: 1995 Body Schema and Intentionality. In J. L. Bermdez, N. Eilan & Marcel. A (Eds.), *The Body and the Self.* Cambridge. Mass.: MIT.

Gallagher, S.: 2001 The Practice of Mind. Theory, Simulation or Primary Interaction? *Journal of Consciousness Studies, 8*(5-7), 83-108.

Gallagher, S.: 2003 Phenomenology and Experimental Design. *Journal of Consciousness Studies, 10*(9-10), 85-99.

Gallagher, S.: 2005 *How the body shapes the mind*. Oxford; New York: Clarendon Press.

Gallagher, S.: 2008a Intersubjectivity in perception. *Continental Philosophy Review, 41*, 163-178.

Gallagher, S.: 2008b Direct perceptions in the intersubjective context. *Consciousness and Cognition, 17*, 535-543.

Gallagher, S.: 2009a Delusional realities. In M. R. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience. Philosophical perspectives*. Oxford: Oxford University Press.

Gallagher, S.: 2009b Philosophical Antecedents. In P. Robbins & M. Aydede (Eds.), *Cambridge Handbook of Situated Cognition*. New York: Cambridge University Press.

Gallagher, S.: 2009c Two Problems of Intersubjectivity. *Journal of Consciousness Studies, 16*(6-8), 289-308.

Gallagher, S.: 2011 Strong Interaction and Self-Agency. *Humana Mente*(15), 55-76.

Gallagher, S., & Hutto, D.: 2008 Understanding others through Primary Interaction and Narrative Practice. In J. Zlatev, T. Racine, C. Sinha & E. Itkonen (Eds.), *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: John Benjamins.

Gallagher, S., & Marcel, A.: 1999a The Self in Contextualized Action. In S. Gallagher & J. Shear (Eds.), *Models of the Self*. Exeter: Imprint Academic.

Gallagher, S., & Marcel, A.: 1999b The Self in Contextualized Action. *Journal of Consciousness Studies, 6*(4), 4-30.

Gandola, M., Invernizzi, P., Sedda, A., Ferrè, E., Sterzi, R., Sberna, M., et al.: 2012 An anatomical account of somatoparaphrenia. *Cortex, 48*(9), 1165-1178

Ganis, G., & Kutas, M.: 2003 An electrophysiological study of scene effects on object identification. *Cognitive Brain Research, 16*, 123-144.

Gardener-Thorpe, & Pearn, J.: 2004 The Cotard syndrome. Report of two patients: with a review of the extended spectrum of 'delire des negations'. *European Journal of Neurology, 11*, 563-566.

Gerrans, P.: 2007 Mechanisms of Madness: evolutionary psychiatry without evolutionary psychology. *Biology and Philosophy, 22*, 35-56.

Gerrans, P.: 2009 Mad scientists or unreliable autobiographers? Dopamine dysregulation and delusions. In M. R. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience*. Oxford: Oxford University Press.

Gerrans, P. & Mulligan, K.: 2012 Imagination and default thinking. Unpublished draft

Ghaemi, S. N.: 2007 Existence and Pluralism. *Psychopathology, 40*, 75-82.

Gilmour, G., Arguello, A., Bari, A., Brown, V., Carter, C., Floresco, S., et al.: 2012 Measuring the construct of executive control in schizophrenia: Defining and validating translational animal paradigms for discovery research. *Neuroscience & Biobehavioral Reviews, (In Press)*.

Goel, V., & Dolan, R.: 2003 Explaining modulation of reasoning by belief. *Cognition,* 87, B11-B22

Gopnik, A., & Meltzoff, A. N.: 1997 *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Graff-Radford, J., Whitwell, J., Geda, Y., & Josephs, K.: 2011 Clinical and imaging features of Othello's syndrome. *European Journal of Neurology, 19*, 38-46.

Greene, M., & Oliva, A.: 2009 The Briefest of Glances. *Psychological Science, 20*(4), 464.

Haney, C., & Zimbardo, P.: 1998 The Past and Future of U.S. Prison Policy. Twenty-five Years After the Stanford Prison Experiment. *American Psychologist, 53*(7), 709-727.

Hebb, D., O.: 1949 *The organization of behavior*. New York: Wiley & Sons.

Heckers, S.: 2001 Neuroimaging Studies of the Hippocampus in Schizophrenia. *Hippocampus, 11*, 520-528.

Heidegger, M.: 2000 *Being and Time* (J. M. a. E. Robinson, Trans. 7 ed.). Oxford: Blackwell.

Hemsley, D. R., & Garety, P.: 1986 The Formation and Maintenance of Delusions: A Bayesian Analysis. *British Journal of Psychiatry, 149*, 51-56.

Hermanowicz, N.: 2002 A Blind Man with Parkinson's Disease, Visual Hallucinations, and Capgras Syndrome. *Journal of Neuropsychiatry and Clinical Neuroscience,* 14(4), 462-463.

Hilgard, E.: 1965 *Hypnotic susceptibility*. New York: Harcourt, Brace & World.

Hilgard, E. R.: 1977 Divided Consciousness: Multiple controls in Human Thought and Action. New York: Wiley.

Hilgard, J.: 1970 Personality and hypnosis: A study of imaginative involvement. Chicago: University of Chicago Press.

Hilgard, J.: 1974 Imaginative Involvement: Some characteristics of the highly hypnotizable and the non-hypnotizable *International Journal of Clinical and Experimental Hypnosis, 22*(2), 138-156.

Hirstein, W., & Ramachandran, V.: 1997 Capgras syndrome: A novel probe for understanding the neural representation of the identity and familiarity of persons. *Proceedings of the Royal Society of London Biological Sciences,* 264(1380), 437-444.

Hohwy, J.: 2004 Top-Down and Bottom-Up in Delusion Formation. *Philosophy, Psychiatry & Psychology,* 11(1), 65-70.

Hohwy, J., & Rajan, V.: 2012 Delusions as Forensically Disturbing Perceptual Inferences. *Neuroethics, 5*, (1) 5-11.

Hohwy, J., & Rosenberg, R.: 2005 Unusual Experiences, Reality Testing and Delusions of Alien Control. *Mind & Language,* 20(2), 141-162.

Huber, M., Karner, M., Kirchler, E., Leppping, P., & Freudenmann, R.: 2008 Striatal lesions in delusional parasitosis revealed by magnetic resonance imaging. *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 32*, 1967-1971.

Huq, S. F., Garety, P., & Hemsley, D. R.: 1988 Probabilistic Judgements in Deluded and Non-Deluded Subjects. *Quarterly Journal of Experimental Psychology, 40A*(4), 801-812.

Husserl, E.: 1981 Phenomenology, Edmund Husserl's article for the *Encyclopaedia Britannica* (1927), in P. McCormick & F. Elliston (Eds.), *Husserl, Shorter works* Notre Dames: University of Notre Dames Press.

Ibanez-Casas, I., & Cervilla, J.: 2012 Neuropsychological Research in Delusional Disorder: A Comprehensive Review. *Psychopathology, 45*, 78-95.

Invernizzi, P., Gandola, M., Romano, D., Zapparoli, L., & Paulescu, E.: 2012 What is mine? Behavioral and anatomical dissociation between somatoparaphrenia and anosognosia for hemiplegia. *Behavioural Neurology, (In Press)*.

Jackson, M., & Fulford, K.: 1997 Spiritual experience and psychopathology *Philosophy, Psychiatry & Psychology, 4*, 41-65.

James, W.: 1890 *The Principles of Psychology* New York: Dover. 1950

Jamieson, G. A., & Woody, E.: 2008 Dissociated control as a paradigm for cognitive neuroscience research and theorizing in hypnosis. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford Handbook of Hypnosis: theory, research and practice*. Oxford UK: Oxford University Press.

Janet, P.: 1901 *The Mental State of Hystericals*. New York: Putnam.

Jaspers, K.: 1968 The Phenomenological Approach in Psychopathology *British Journal of Psychiatry, 114*, 1313-1323.

Jaspers, K.: 1997 (1913) *General Psychopathology* (J. Hoenig & M. Hamilton, Trans. 1963) Baltimore: John Hopkins University Press.

Jolley, S., Garety, P., Bebbington, P., dunn, G., Freeman, D., Kuipers, E., et al.: 2006 Attributional style in psychosis - The role of affect and belief type. *Behaviour Research and Therapy, 44*, 1597 - 1607.

Johnson, M. K.: 1991 Reality Monitoring: Evidence from Confabulation in Organic Brain Disease Patients In G. Prigantano & D. Schacter (Eds.), *Awareness of Deficit After Brain Injury. Clinical and Theoretical Issues*. New York: Oxford University Press.

Johnson, M., K, Hashtrudi, S., & Lindsay, D.: 1993 Source Monitoring. *Psychological Bulletin, 114*(1), 3-28.

Jones, A.: 1966 Information deprivation in humans. In B. A. Maher (Ed.), *Progress in experimental personality research* Vol. 3. New York: Academic Press.

Josephs, K.: 2007 Capgras Syndrome and Its Relationship to Neurodegenerative Disease. *Archives of Neurology,* 64(12), 1762-1766.

Joubert, O., Rousselet, G., Fize, D., & Fabre-Thorpe, M.: 2007 Processing scene context: Fast categorization and object interference. *Vision Research 47*, 3286-3297.

Joubert, O., Fize, D., Rousselet, G., & Fabre-Thorpe, M.: 2008 Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision, 8*(13:11), 1-18.

Kaney, S & Bentall, R. P.: 1989 Persecutory delusions and attributional style *British Journal of Medical Psychology 63*, 191-198

Kapur, S.: 2003 Psychosis as a State of Aberrant Salience: A Framework Linking Biology, Phenomenology, and Pharmacology in Schizophrenia. *American Journal of Psychiatry, 160*, 13-23.

Kendler, K., & Schaffner, K.: 2011 The Dopamine Hypothesis of Schizophrenia: An Historical and Philosophical Analysis. *Philosophy, Psychiatry & Psychology, 18*(1), 41-63.

Kihlstrom, J. F.: 2008 The domain of hypnosis revisited. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford Handbook of Hypnosis: Theory, research and practice*. Oxford UK: Oxford University Press.

Kimura, B.: 2001 *Cogito* And I: A Bio-Logical Approach. *Philosophy, Psychiatry & Psychology, 8*(4), 331-336.

Kirsch, I.: 2011 The Altered State Issue: Dead or Alive? *International Journal of Clinical and Experimental Hypnosis,* 59(3), 350-362.

Kirsch, I., & Lynn, S.: 1995 Altered states of hypnosis: Changes in the theoretical landscape. *American Psychologist, 50*(10), 846-858.

Klee, R.: 2004 Why Some Delusions Are Necessarily Inexplicable Beliefs. *Philosophy, Psychiatry & Psychology, 11*(1), 25-34.

Knill, D., & Pouget, A.: 2004 The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Cognitive Science, 27*(12), 712-719.

Kompus, K., Westerhausen, R., & Hugdahl, K.: 2011 The "paradoxical" engagement of the primary auditory cortex in patients with auditory verbal hallucinations: A met-analysis of functional neuroimaging. *Neuropsychologia, 40*, 3361-3369.

Kravitz, D., Peng, C., & Baker, C.: 2011 Real-World Scene Representations in High-Level Visual Cortex: It's the Spaces More Than the Places. *Journal of Neuroscience, 31*(20), 7322-7333.

Krishnan, K. R. R.: 1990 Brain imaging and psychiatric disorders. *Current Opinion in Psychiatry, 3*, 79-82.

Lamme, V.: 2006 Towards a true neural stance on consciousness. *Trends in Cognitive Science, 10*(11), 494-501.

Langdon, R., & Bayne, T.: 2010 Delusion and confabulation: Mistakes of perceiving, remembering and believing. *Cognitive Neuropsychiatry, 15*(1/2/3), 319-345.

Langdon, R., & Coltheart, M.: 2000 The Cognitive Neuropsychology of Delusions. *Mind & Language, 15*(1), 184-218.

Laurence, S., & Margolis, E.: 1999 Concepts and Cognitive Science. In E. Margolis & S. Laurence (Eds.), *Concepts. Core Readings*. London: MIT.

Leboe, J., & Whittlesea, B.: 2002 The Inferential Basis of Familiarity and Recall: Evidence for a Common Underlying Process. *Journal of Memory and Language,* 46, 804-829.

Lewis, M. B., Sherwood, S., Moselhy, H., & Ellis, H.: 2001 Autonomic responses to familiar faces without autonomic responses to familiar voices: Evidence for voice-specific Capgras delusion. *Cognitive Neuropsychiatry,* 6(3), 217-228.

Lopez, C., Lenggenhager, B., & Blanke, O.: 2010 How vestibular stimulation interacts with illusory hand ownership. *Consciousness and Cognition, 19*(1), 33-47.

Lopez, C., Biere, C., Preuss, N., & Mast, F.: 2012 Tactile and vestibular mechanisms underlying ownership for body parts: A non-visual variant of the rubber hand illusion. *Neuroscience Letters, 511*, 120-124.

Lucchelli, F., & Spinnler, H.: 2008 A reappraisal of person recognition and identification. *Cortex, 44*, 230-237.

Luhrmann, T.: 2011 Hallucinations and Sensory Overrides. *Annual Review of Anthropology, 40*, 71-85.

Lykouras, L., Typaldou, M., Gournellis, R., Vaslamatzis, G., & Christodoulou, G.-N.: 2002 Coexistence of Capgras and Fregoli syndromes in a single patient. Clinical, neuroimaging and neuropsychological findings. *European Psychiatry, 17*, 234-235.

Lynn, S., & Green, J.: 2011 The Sociocognitive and Dissociation Theories of Hypnosis: Towards a Rapprochement. *International Journal of Clinical and Experimental Hypnosis, 59*(3), 277-293.

Machery, E.: 2010 Précis of *Doing without Concepts*. *Mind & Language, 25*(5), 602-611.

Maguire, A. M., & Ogden, J. A.: 2002 MRI brain scan analysis and neuropsychological profiles of nine patients with persisting unilateral neglect. *Neuropsychologia,* 40, 879-887

Maher, B.: 1974 Delusional Thinking and Perceptual Disorders. *Journal of Individual Psychology, 30/1,* 98-113.

Maher, B.: 1988a Anomalous Experience and Delusional Thinking: The Logic of Explanations. In T. Oltman & B. Maher (Eds.), *Delusional Beliefs* New York: John Wiley & Sons.

Maher, B.: 1988b Delusions as the Product of Normal Cognition. In T. Oltmanns & B. Maher (Eds.), *Delusional Beliefs* New York: John Wiley & Sons.

Maher, B.: 1990 The irrelevance of Rationality to Adaptive Behavior. In M. Spitzer & B. Maher (Eds.) *Philosophy and Psychopathology* New York: Springer-Verlag.

Maher, B.: 1999 Anomalous Experiences in Everyday Life: Its Significance for Psychopathology *Monist, 82/4*, 547-571.

Maher, B.: 2003a Psychopathy and Delusions: Reflections on Methods and Models. In M. Lenzenweger & J. Hooley (Eds.), *Principles of experimental psychopathology: essays in honor of Brendan Maher* Washington DC: American Psychological Association

Maher, B.: 2003b Schizophrenia, Aberrant Utterances and Delusions of Control: The Disconnection of Speech and Thought, and the Connection of Experience and Belief. *Mind & Language, 18*(1), 1-22.

McDermott, D.: 2007 Dodging the explanatory gap - or bridging it. *Behavioral and Brain Sciences, 30*(5/6) 518-518.

Maloney, L., & Mamassian, P.: 2009 Bayesian decision theory as a model of visual perception: Testing Bayesian transfer. *Visual Neuroscience, 26*, 147-155.

Maloney, L., & Zhang, H.: 2010 Decision-theoretic models of visual perception and action. *Vision Research, 50*, 2362-2374.

Marcel, A.: 1992 The personal level in cognitive rehabilitation. In N. Von Steinbüchel, E. Pöppel and D. Cramon (Eds) *Neuropsychological Rehabilitation.* Berlin. Springer

Marcel, A.: 1993 Slippage in the unity of consciousness. In *Experimental and Theoretical Studies of Consciousness*, Ciba Foundation Symposium #174. Chichester: John Wiley & Sons

McAllister, T., & Ferrell, R.: 2002 Evaluation and treatment of psychosis after traumatic brain injury. *NeuroRehabilitation, 17*, 357-368.

McCauley, R., & Henrich, J.: 2006 Susceptibility to the Müller-Lyer Illusion, Theory-Neutral Observation, and the Diachronic Penetrability of the Visual Input System. *Philosophical Psychology, 19*(1), 1-23.

McConkey, K. M., Szeps, A., & Barnier, A. J.: 2001 Indexing the experience of sex change in hypnosis and imagination. *International Journal of Clinical and Experimental Hypnosis, 49*, 123-138.

McKay, R., Langdon, R., & Coltheart, M.: 2005 Paranoia, persecutory delusions and attributional biases. *Psychiatry Research, 136*, 233-245.

McKay, R., Langdon, R., & Coltheart, M.: 2007 Jumping to delusions? Paranoia, probabilistic reasoning and need for closure. *Cognitive Neuropsychiatry, 12*(4), 362-376.

Menon, M., Schmitz, T., Anderson, A., Graff, A., Korostil, M., Mamo, D., et al.: 2011 Exploring the Neural Correlates of Delusions of Reference. *Biological Psychiatry, 70*, 1127-1133.

Mentis, M., Weinstein, E., Horwitz, B., McIntosh, A., Pietrini, P., Alexander, G., et al.: 1995 Abnormal Brain Glucose Metabolism in the Delusional Misidentification Syndromes: A Positron Emission Tomography Study in Alzheimer Disease. *Biological Psychiatry, 38*, 438-449.

Menzies, V., Taylor, A., & Bourguignon, C.: 2008 An individual Difference to Consider in Mind-Body intervention. *Journal of Holistic Nursing, 26*(4), 297-302.

Meredith, S., Whyler, N., Stanford, A., Chakirova, G., Moorehead, T., Job, D., et al.: 2012 Anterior cingulate morphology in people at high-risk of schizophrenia. *European Psychiatry, 27*(5), 377-385.

Miller, A.: 2011 The Development of the Theories of Meaning: From Frege to McDowell and Beyond. In M. Beaney (Ed.), *The Oxford Handbook of the History of Analytic Philosophy*. Oxford: Oxford University Press. *(In Press)*

Minkowski, E.: 1926 Bergson's Conceptions As Applied To Psychopathology. *Journal of Nervous & Mental Disease, 63*(6), 553-568.

Minkowski, E.: 1927 Schizophrenia. Psychopathology of schizoids and schizophrenics. Oxford: Payot.

Minkowski, E., & Targowla, R.: 1923 (Ziadeh trans.) 2001 A Contribution To The Study Of Autism: The Interrogative Attitude. *Philosophy, Psychiatry & Psychology, 8*(4), 271-278.

Mitchell, K. J., & Johnson, M. K.: 2000 Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory*. New York: Oxford University Press.

Moore, J., & Fletcher, P. C.: 2012 Sense of agency in health and disease: A review of cue integration processes. *Consciousness and Cognition, 21*, 59-68.

Mori, H.: 2010 A human fetus development simulation: Self-organization of behaviors through tactile sensation. *Proceedings of the 9th IEEE 2010 International Conference on Development and Learning*, 82-87.

Moritz, S., & Wooward, T.: 2006 A generalized bias against disconfirmatory evidence in schizophrenia. *Psychiatry Research, 142*, 157-165.

Moriyama, Y., Muramatsu, T., Kato, M., Mimura, M., Akiyama, T., & Kashima, H.: 2007 Fregoli Syndrome Accompanied with Prosopagnosia in a Woman with a 40-year History of Schizophrenia. *The Keio Journal of Medicine, 56*(4), 139-134.

Morrison, A., Haddock, G., & Tarrier, N.: 1995 Intrusive thoughts and auditory hallucinations: A cognitive approach. *Behavioural and Cognitive Psychotherapy, 23*(3), 265-280.

Mullen, P.: 2007 A Modest Proposal for Another Phenomenological Approach to Psychopathology. *Schizophrenia Bulletin, 33*(1), 113-121.

Mullins, S., & Spence, S. A.: 2003 Re-examining thought insertion. *British Journal of Psychiatry, 182*, 293 -298.

Murray, G. K., Corlett, P. R., & Fletcher, P. C.: 2010 The Neural Underpinnings of Associative Learning in Health and Psychosis: How Can Performance Be Preserved When Brain Responses Are Abnormal. *Schizophrenia Bulletin, 36*(3), 465-471.

Nagaratnam, N., & O'Neile, L.: 2000 Delusional Parasitosis Following Occipito-Temporal Cerebral Infarction. *General Hospital Psychiatry, 22*(2), 129-132.

Nejad, A. G., & Toofani, K.: 2006 A variant of Capgras syndrome with delusional conviction of inanimate doubles in a patient with grandmal epilepsy. *Acta Neuropsychiatrica* 18, 52-54.

Nichols, S.: 2008 Imagination and the *I*. *Mind & Language, 23*(5), 518-535.

Nordahl, T., Carter, C., Salo, R., Kraft, L., Baldo, J., Salamat, S., et al.: 2001 Anterior Cingulate Metabolism Correlates with Stroop Errors in Paranoid Schizophrenia. *Neuropsychopharmacology, 25*(1), 139-148.

Oakley, D., & Halligan, P.: 2009 Hypnotic suggestion and cognitive neuroscience. *Trends in Cognitive Science, 13*(6), 264-270.

O'Carroll, R. E., & Prentice, N.: 1996 Functional adaptation to traumatic brain injury; recovery of function following left hemisphere damage acquired in adulthood. *Neurocase, 2*, 503-308.

Okubo, Y., Sekiya, H., Namiki, S., Sakamoto, H., Iinuma, S., Yamasaki, M., et al.: 2010 Imaging extrasynaptic glutamate dynamics in the brain. *PNAS, 107*(14), 6526-6531.

Oliva, A.: 2005 Gist of the Scene. In I. Laurent, R. Geraint & T. John (Eds.), *Neurobiology of Attention*. Amsterdam; Boston: Elsevier Academic Press.

Olojugba, C., de Silva, R., Kartsounis, L., Royan, L., & Carter, J.: 2007 De Clerambault's syndrome (erotomania) as a presenting feature of fronto-temporal dementia and motor neurone disease. *Behavioural Neurology, 18*, 193-195.

Oppenheim, G., & Dell, G.: 2008 Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition, 106*(1), 528-537.

Pacherie, E.: 2009 Perception, Emotions, and Delusions: The Case of the Capgras Delusion. In T. Bayne and J. Fernández (eds.), *Delusion and Self-Deception. Affective and Motivational Influences on Belief Formation*. New York: Psychology Press.

Palmer, S. E.: 1975 The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3*, 519-526.

Parnas, J., & Sass, L. A.: 2001 Self, Solipsism, And Schizophrenic Delusions. *Philosophy, Psychiatry & Psychology, 8*(2-3), 1001-1120.

Parnas, J., Sass, L. A., & Zahavi, D.: 2008 Recent developments in philosophy of psychology. *Current Opinion in Psychiatry, 21*, 278-584.

Pearl, M., Talgat, G., Valea, F., & Chalas, E.: 1998 Psychiatric Symptoms Due to Brain Metastases. *Medical update for psychiatrists, 3*(4), 91.

Poldrack, R., & Wagner, A.: 2004 What Can Neuroimaging Tell Us About the Mind? Insights From Prefrontal Cortex. *Current Directions In Psychological Science, 13*(5), 177-181.

Potter, M. C.: 1975 Meaning in visual search. *Science, 187*, 965-966.

Potter, M. C.: 1976 Short-term conceptual memory for pictures. *Journal of Experimental Psychology Human Learning and Memory, 2*, 509-522.

Price, B., & Mesulam, M.: 1985 Psychiatric Manifestations of Right Hemisphere Infarctions. *Journal of Nervous & Mental Disease, 173*(10), 610-614.

Pun, B., Lekh, S., Nijran, K., Bagary, M., & Richardson, A.: 2011 SPECT neuroimaging in schizophrenia with religious delusions. *International Journal of Psychophysiology, 40*(2), 143-148.

Radden, J.: 2011 *On Delusions: Thinking in Action*. London: Routledge.

Raffard, S., & Bayard, S.: 2012 Understanding the executive functioning heterogeneity in schizophrenia. *Brain and Cognition, 79*(1), 60-69.

Raftopoulos, A.: 2001 Is perception informationally encapsulated? The issue of the theory-ladeness of perception. *Cognitive Science, 25*(3), 423-451.

Raftopolous, A.: 2009 Cognition and Perception: How do Psychology and Neural Science Inform Philosophy? Massachusetts: MIT Press.

Rahmanovic, A., Barnier, A. J., Cox, R., Langdon, R., & Coltheart, M.: 2012 "That's not my arm": A hypnotic analogue of somatoparaphrenia. *Cognitive Neuropsychiatry, 17*(1), 36-63.

Ratcliffe, M.: 2007 What Is a Feeling of Unfamiliarity? *Philosophy, Psychiatry & Psychology,* 14(1), 43-49.

Ratcliffe, M.: 2010 The phenomenology of mood and the meaning of life. In P. Goldie (Eds.), *The Oxford Handbook of Philosophy of Emotion. Oxford UK: Oxford University Press.*

Raz, A.: 2011 Does Neuroimaging of Suggestion Elucidate Hypnotic Trance? *International Journal of Clinical and Experimental Hypnosis, 59*(3), 363-377.

Reid, I., Young, A. W., and Hellawell, D. J.: 1993 Voice recognition impairment in a blind Capgras patient. *Behavioural Neurology 6*(4), 225-228

Reimer, M.: 2009 Is the impostor hypothesis really so preposterous? Understanding the Capgras experience. *Philosophical Psychology, 22*(6), 669-686.

Rodlier, M., Prevost, M., Renoult, L., Lionnet, C., Kwann, Y., Dionne-Dostie, E., et al.: 2011 Healthy people with delusional ideation change their mind with conviction. *Psychiatry Research, 189*, 433-439.

Rojo, V., Caballero, L., Iruela, L., & Baca, E.: 1991 Capgras Syndrome in a Blind Patient. *American Journal of Psychiatry,* 148(9), 1271-1272.

Rosler, A., Holder, G., & Seifritz, E.: 2001 Canary Capgras. *Journal of Neuropsychiatry and Clinical Neuroscience,* 13(3), 429.

Rothlisberger, M., Riecher-Rossier, A., Aston, J., Fusar-Poli, P., Radu, E.-W., & Borgwardt, S.: 2012 Cingulate Volume Abnormalities in Emerging Psychosis. *Current Pharmaceutical Design, 18*(4), 495-504.

Sadler, P., & Woody, E.: 2010 Dissociation in Hypnosis: Theoretical Frameworks and Psychotherapeutic Implications. In S. Lynn, J. Rhue & I. Kirsch (Eds.), *Handbook of Clinical Hypnosis* (2 ed.). Washington: American Psychological Assoc.

Sass, L. A.: 1992a Madness and Modernism. Insanity in the Light of Modern Art, Literature and Thought. New York: Harper Collins.

Sass, L. A.: 1992b Heidegger, schizophrenia and the ontological difference. *Philosophical Psychology, 5*(2), 109-132.

Sass, L. A.: 1994 The Paradoxes of Delusion. Wittgenstein, Schreber, and the Schizophrenic Mind. New York: Cornell University.

Sass, L. A.: 1998 Interpreting Schizophrenia: Construal or Construction? A Reply to Robert J. Barrett. *Culture, Medicine and Psychiatry, 22*, 495-503.

Sass, L. A.: 1999 Analyzing and Deconstructing Psychopathology. *Theory & Psychology, 9*(2), 257-268.

Sass, L. A.: 2001 Self And World In Schizophrenia: Three Classic Approaches. *Philosophy, Psychiatry & Psychology, 8*(4), 251-269.

Sass, L. A.: 2004 Affectivity in Schizophrenia: A Phenomenological View. *Journal of Consciousness Studies, 11*(10-11), 127-147.

Sass, L. A.: 2007 'Schizophrenic Person' or 'Person with Schizophrenia'? *Theory & Psychology, 17*(3), 395-420.

Sass, L. A., & Parnas, J.: 2001 Phenomenology Of Self-Disturbances In Schizophrenia: Some Research Findings And Directions. *Philosophy, Psychiatry & Psychology, 8*(4), 347-356.

Sass, L. A., & Parnas, J.: 2003 Schizophrenia, Consciousness, and the Self. *Schizophrenia Bulletin, 29*(3), 427-444.

Sass, L. A., Parnas, J., & Zahavi, D.: 2011 Phenomenological Psychopathology and Schizophrenia: Contemporary Approaches and Misunderstandings. *Philosophy, Psychiatry & Psychology, 18*(1), 2-23.

Schutz, A.: 1974 *Collected Papers Vol. 1. The Problem of Social Reality.* Dordrecht: Springer

Sellal, F., F, F. S., Van Der Linden, M., Rainville, C., & Labrecque, R.: 1996 To Be or Not to Be at Home? A Neuropsychological Approach to Delusion for Place. *Journal of Clinical and Experimental Neuropsychology, 18*(2), 234-248.

Silva, J. A., Leong, G. B., Weinstock, R., Sharma, K.et al.: 1994 Delusional misidentification syndromes and dangerousness. *Psychopathology, 27*(3-5), 215-219.

Sousa, P., & Swiney, L.: 2011 Thought insertion: Abnormal sense of thought agency or thought endorsement. *Phenomenology and the Cognitive Sciences*, *(In Press)*

Southard, E. E.: 1912 On the somatic sources of somatic delusions *Journal of Abnormal Psychology, 7,* 326-339.

Southard, E. E.: 1913 The possible correlation between delusions and cortex lesions in general apresis *Journal of Abnormal Psychology, 8,* 259-275.

Southard, E. E.: 1915 Data concerning delusions of personality with a note on the association of Bright's disease and unpleasant delusions *Journal of Abnormal Psychology, 10,* 241-262.

Southard, E. E.: 1916 On descriptive analysis of manifest delusions from the subject's point of view *Journal of Abnormal Psychology, 11,* 189-202.

Spence, S. A.: 2001 Alien Control: From Phenomenology to Cognitive Neurobiology. *Philosophy, Psychiatry & Psychology, 8*(2/3), 163-172.

Spitzer, M.: 1990 Why Philosophy? In M. Spitzer & B. Maher (Eds.), *Philosophy and Psychopathology*. New York: Springer-Verlag.

Startup, H., Freeman, D., & Garety, P.: 2008 Jumping to conclusions and persecutory delusions. *European Psychiatry,* 23, 457-459.

Stephens, G. & Graham, G.: 1994 Self-consciousness, mental agency and the clinic psychopathology of thought insertion *Philosophy, Psychiatry & Psychology 1* 1-10.

Stone, T., & Young, A. W.: 1997 Delusions and Brain Injury: The Philosophy and Psychology of Belief. *Mind & Language, 12*(3/4), 327-364.

Sun, H.-M., Simon-Dack, S., Gordon, R., & Teder, W.: 2011 Contextual influences on rapid object categorization in natural scenes. *Brain Research, 1398*, 40-54.

Sutcliffe, J. F.: 1960 "Credulous" and "Skeptical" Views of Hypnotic Phenomena. *Journal of Abnormal and Social Psychology, 62*(2), 189-200.

Taber, K., & Hurley, R.: 2007 Neuroimaging in Schizophrenia: Misattributions and Religious Delusions. *The Journal of Neuropsychiatry and Clinical Neuroscience, 19*(iv-4).

Talley, B., & Michels, L.: 2009 Two Patients who Think their Family Members have been Replaced by 'Imposters'. *Psychiatric Annals, 39*(5), 247-253.

Tamam, L., Karatas, G., Zeren, T., & Ozpoyraz, N.: 2003 The prevalence of Capgras syndrome in a university hospital setting. *Acta Neuropsychiatrica,* 15, 290-295.

Tausk, V.: 1988 On the origin of the 'influencing machine' in schizophrenia In P. Buckley (Ed.), *Essential Papers on Psychosis*. New York: New York University Press.

Tellegen, A., & Gilbert, A.: 1974 Openness to Absorbing and Self-Altering Experiences ("Absorption"), a Trait Related to Hypnotic Susceptibility. *Journal of Abnormal Psychology, 83*(3), 268-277.

Todd, J., Dewhurst, K., & Wallis, G.: 1981 The Syndrome of Capgras. *British Journal of Psychiatry,* 139, 319-327.

Tost, H., Ruf, M., Schmail, C., Schultze, T., Knorr, C., Vollmert, C., et al.: 2010 Prefrontal-temporal gray matter deficits in bipolar disorder patients with persecutory delusions. *Journal of Affective Disorders, 120*(1-3), 54-61.

Tranel, D., Damasio, H., & Damasio, A.: 1995 Double Dissociation between Overt and Covert Face Recognition. *Journal of Cognitive Neuroscience, 7*(4), 425-432.

Trevarthen, C. B.: 1979 Secondary Intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A Lock (Ed.), *Action Gesture, and Symbol: The Emergence of Language* London: Academic Press

Tsujino, N., Nemoto, T., Yamaguchi, T., Katagiri, N., Tohgi, N., Ikedo, R., et al.: 2011 Cerebral blood flow changes in very-late-onset schizophrenia-like psychosis with catatonia before and after successful treatment. *Psychiatry and Clinical Neuroscience, 65*(6), 600-603.

Vallar, G., & Ronchi, R.: 2009 Somatoparaphrenia: a body delusion. A review of the neuropsychological literature. *Experimental Brain Research, 192*, 533-551.

Van den Stock, J., de Gelder, B., De Winter, F.-L., Van Laere, K., & Vandenbulcke, M.: 2012 A strange face in the mirror. Face-selective self-misidentification in a patient with right lateralized occipito-temporal hypo-metabolism. *Cortex, (In Press)*.

van Veen, V., & Carter, C.: 2002 The anterior cingulate as a conflict monitor: fRMI and ERP studies. *Physiology & Behavior, 77*, 477-482.

Verkhratsky, A., & Kirchhoff, F.: 2007 NMDA Receptors in Glia. *Neuroscience, 13*, 28-37.

Villarejo, A., Martin, V., Moreno-Ramos, T., Camacho-Salas, A., Porta-Etessam, J., & Bermejo-Pareja, F.: 2011 Mirrored-self misidentification in a patient without dementia: evidence for right hemispheric and bifrontal damage. *Neurocase: The Neural Basis of Cognition, 17*(3), 276-284.

Vinogradova, O. S., Willis-Shore, J., Poole, J., Marten, E., Ober, B., & Shenaut, G.: 1997 Clinical and Neurocognitive Aspects of Source Monitoring Errors in Schizophrenia *American Journal of Psychiatry, 154* (11), 153-1537.

Walker, C. 1988 Philosophical concepts and practice: the legacy of Karl Jaspers' psychopathology. *Current Opinion in Psychiatry*, 1, 624-629.

Walker, C.: 1991 Delusion: What Did Jaspers Really Say? *British Journal of Psychiatry, 159* (suppl. 1 4), 94-103.

Wallis, G.: 1986 Nature of the Misidentified in the Capgras Syndrome. *Bibliotheca Psychiatrica, 164*, 40-48.

Walton, K.: 1978 How Remote are Fictional Worlds from the Real World? *J. of Aesthetics and Art Criticism, 37*(1), 11-23.

Walton, K.: 1990 *Mimesis as Make-Believe. On the Foundations of the Representational Arts*. Massachusetts: Harvard University Press.

Walton, K.: 1991 Précis of Mimesis as Make-Believe: On the Foundations of the Representational Arts. *Philosophy and Phenomenological Research, 11*(2).
Weisberg, D., & Bloom, P.: 2009 Young children separate multiple pretend worlds *Developmental Science, 12*(5), 699-705.
Weitzenhoffer, A., & Hilgard, E.: 1959 *Stanford Hypnotic Susceptibility Scales, Forms A & B*. Palo Alto CA: Consulting Psychologists Press.
Weitzenhoffer, A. M. & Hilgard, E. R.: 1962 Stanford Hypnotic Susceptibility Scale Form C. Modified by John F. Kihlstrom: Stanford University
Wellman, H. M., & Estes, D.: 1986 Early understanding of mental entities: a reexamination of childhood realism. *Child Development, 57*, 910-923.
Westerhausen, R., Kompus, K., & Hugdahl, K.: 2011 Impaired cognitive inhibition in schizophrenia: A meta-analysis of the Stroop interference effect. *Schizophrenia Research, 133*, 172-181.
Whitehead, D., Tunnard, C., Hurt, C., Wahlund, L., Mecocci, P., Tsolaki, M., et al.: 2012 Frontotemporal atrophy associated with paranoid delusions in women with Alzheimers disease. *International Psychogeriatrics, 24*(1) 99-107.
Whitford, T., Mathalon, D., Shenton, M., Roach, B., Bammer, R., Adcock, A., et al.: 2011 Electrophysiological and diffusion tensor imaging evidence of delayed corollary discharges in patients with schizophrenia. *Psychological Medicine, 41*, 959-969.
Whittlesea, B., & Leboe, J.: 2003 Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language,* 49, 62-79.
Whittlesea, B., & Williams, L.: 2000 The Source of Feelings of Familiarity: The Discrepancy-Attribution Hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 26(3), 547-565.
Whittlesea, B., & Williams, L.: 2001 The Discrepancy-Attribution Hypothesis: II. Expectation, Uncertainty, Surprise, and Feelings of Familiarity. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 27(1), 14-33.
Winters, K., & Neale, J.: 1983 Delusions and Delusional Thinking. *Clinical Psychology Review, 3*, 227-253.
Wittgenstein, L.: 1972 *On Certainty.* Paul Denis & G. E. M. Anscombe (Trans.) New York: Harper & Row.

Wong, A., & Meier, R.: 1997 Case Report: Delusional Jealousy Following Right-Sided Cerebral Infarct. *Neurocase, 3*, 391-394.

Woody, E., & Sadler, P.: 2008 Dissociation theories of hypnosis. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford Handbook of Hypnosis: theory, research and practice*. Oxford UK: Oxford University Press.

Woolley, J. D., & Phelps, K. E.: 1994 Young children's practical reasoning about imagination. *British Journal of Developmental Psychology, 12*, 53-67.

Yeung, N., Cohen, J., & Botvinick, M.: 2011 Errors of interpretation and modeling: A reply to Grinband et al. *Neuroimage, 57*, 316-319.

Young, A.: 1992 Face recognition impairments. *Philosophical Transactions: Biological Sciences, 335*(1273), 47-53.

Young, A.: 1999 Delusions. *Monist, 82*(4), 571-589.

Young, A.: 2000 Wondrous Strange: The Neuropsychology of Abnormal Beliefs. *Mind & Language, 15*(1), 47-73.

Young, A., Ellis, H., & Szulecka, T. K.: 1990 Face processing impairments and delusional misidentification. *Behavioural Neurology, 3*, 153-168.

Young, G.: 2008 Capgras Delusion: An interactionist model. *Consciousness and Cognition, 17*(3), 863-876.

Young, H., & Bentall, R.: 1997 Probabilistic reasoning in deluded, depressed and normal subjects: effects of task difficulty and meaningful *versus* non-meaningful material. *Psychological Medicine, 27*, 455-465.

Zimbardo, P.: 1973 On the ethics of intervention in human psychological research: With special reference to the Stanford prison experiment. *Cognition, 2*(2), 243-256.

Zimbardo, P.: 2007 Revisiting the Stanford Prison Experiment: A Lesson in the Power of Situation. *The Chronicle Review, 53*(30), B6.

Zubeck, J. P., Sansom, W., & Phrsiazniuk, A.: 1960 Intellectual changes during prolonged sensory isolation (darkness and silence). *Canadian Journal of Psychology 14,* 83-100

Zubeck, J. P., Aftansa, M., Hasek, J., Sansom, W., Schludermann, E., Wilgosh, L., & Winocur, G.: 1962 Intellectual and perceptual changes during prolonged

perceptual deprivation: low illumination and noise level *Perceptual Motor Skills 15,* 171-198