A Dennettian plus Prediction Error Account of Delusion

Deborah Wells

PhD

University of York

Philosophy

April 2022
ABSTRACT

The thesis approaches the problem of delusion as delusion occurs in the clinical field, and identifies two questions to be addressed. The first question is: which account of delusion most comprehensively addresses the dataset of clinical features of delusion? The second question asks: is this successful account an empiricist or a rationalist or a different approach?

Three delusions are chosen as test cases because together they represent a wide range of the data set of clinical features. The delusions are: erotomaniac delusion; Capgras delusion; and perceptual delusional bicephaly.

First the prevailing theories of delusion are examined. These are: the empiricist one-factor approach; the empiricist two-factor approach; and the (relationist) rationalist approach to delusion. The conclusion is that in virtue of stipulations which each of the prevailing theories variously makes concerning neuropathology; (deep) irrationality; and abnormal experience, all of the accounts encounter substantial problems. I use this conclusion as warrant to propose an alternative account.

The alternative account which I develop combines two discrete theories. The theories are: Daniel Dennett’s approach to mind and Prediction Error theory as applied to the brain. First I build up a combined Dennettian plus Prediction Error (D-PE) model. This is then applied to the test cases of delusion, where it is shown to comprehensively meet the features of each case.

Therefore to answer the first thesis question, a D-PE approach affords the greatest explanatory potential. And to answer the second thesis question, while a D-PE account is technically a rationalist approach to delusion, it is different because, pace rationalism as typically construed, the D-PE model eschews the idea of (deep) irrationality and it permits abnormal experience an explanatory role. Therefore a D-PE approach adds to the ongoing conversation by offering a different way of thinking about the problem of delusion in the clinical field.
# Table of Contents

Abstract 2

Table of Contents 3

Acknowledgement 5

Author's Declaration 6

Chapter 1: The Problem of Delusion 7
  Introduction: The Problem of Delusion 7
  Features of Delusion: Three Case Studies 11
  The Prevailing Approaches to Explaining Delusion 24
  A Dennettian plus Prediction Error Approach 32
    Daniel Dennett’s Theory of Mind 33
    Prediction Error Theory 47
  Implications of Combining the Dennettian and Prediction Error Theses 51
  The Organization of the Thesis 55

Chapter 2: The Prevailing Approaches to Explaining Delusion 57
  Introduction 57
  The Three Prevailing Theories of Delusion 58
    The Empiricist One-Factor Theory 58
    The Empiricist Two-Factor Theory 61
    The (Relationist) Rationalist Theory 63
  Appeal to Neuropathology 69
  Appeal to (Deep) Irrationality 74
  Appeal to Abnormal Experience 91

Chapter 3: A Dennettian plus Prediction Error Account 122
  Introduction: Moving Towards a Combined Account 122
  The Normal Case Study of Dennett and the Molly B 129
  A Dennettian plus Prediction Error Framework 130
How Dennett Makes Up then (Perhaps) Changes His Mind 155
When First Dennett Sees the Molly B 163

Chapter 4: A Dennettian plus Prediction Error Account of Delusion 168

Introduction 168
Three Cases of Delusion 169
   Erotomantic Delusion 169
   Capgras Delusion 199
   Perceptual Delusional Bicephaly 207
Concluding Considerations 216

References 219
I would like to acknowledge
and observe the debt of gratitude which I owe to
Professor Daniel C. Dennett.

Thank you.
AUTHOR DECLARATION

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references.
CHAPTER 1

THE PROBLEM OF DELUSION

Introduction: The Problem of Delusion

The aim of this thesis is to explore as a philosophical problem the concept of delusion as delusion occurs in cases of psychosis in the clinical field. This is not to suggest that delusion is something which only ever relates to a clinical environment. For instance, one might subscribe to a continuum view of psychosis, which is the view that delusion-type thoughts exist on the same continuum as commonplace thoughts (e.g. Johns and van Os 2002; Verdoux and van Os 2002; Rossler et al. 2007); whereas in philosophy delusion is often approached as an abstraction, with the emphasis not so much on a subject (patient) who suffers a delusion, so much as on the concept of a fixed false belief (e.g. Maher 1974; Coltheart et al. 2011; Campbell 2001). However, here my approach addresses the concept of a delusion, but in a context which very much includes the person who suffers delusion and the special or difficult circumstances which he or she has to negotiate in his or her life – which may or may not involve an organic pathology of the brain – and which (circumstances) I argue may have an important role in that person developing a clinical delusion, which in line with the continuum hypothesis, I take to present as part of the same continuum as the non-clinical convictions the person holds.

Terminology and concepts

To clarify the terminology and concepts I use. Starting with the concepts of psychosis and psychotic disorder as these establish the field in which clinical delusions occur. Essentially, psychosis is a severe state of mental disorder in which a subject is in some way divorced from what we might loosely term ‘objective reality’ or more accurately from certain expected norms, and where the condition of disorder is met by the occurrence of particular mental and physical characteristics (i.e. psychotic features or symptoms) which together serve to indicate that the subject has a problem which is of a type and serious enough to warrant clinical (psychiatric) concern (cf. APA 2013 pp. 87-90). In contrast, a psychotic disorder is the specific psychiatric condition with which a patient is diagnosed on the basis of the pattern of psychotic features with which he or she presents, and following which diagnosis a pertinent course of treatment proceeds
Examples of psychotic disorders are delusional disorder (pp. 90-3) and schizophrenia (pp. 99-105), the first of which necessarily entails the occurrence of delusion (p. 90), and the second of which lists delusion as one of three individually sufficient but not individually necessary features in the overall pattern of diagnostic criteria (p. 99). Both of these psychoses feature in the case studies of delusion which follow.

Concerning the characteristic features of psychosis, they fall into five groups. The groups are: Delusions, i.e. maintaining a conviction about something which is manifestly at odds with recognised (cultural, social, logical) norms; Hallucinations, i.e. a perceptual-type experience which occurs in the absence of a significant perceptual stimulus; Significantly disordered thoughts, or more correctly speech, e.g. frequent derailments or marked disorganisation; Significantly abnormal (motor) behaviours, e.g. obsessive behaviours or behavioural tics; Negative symptoms, i.e. a marked absence of normal behaviours, such as an absence of motivation or emotion, or a failure by the subject to maintain his or her basic needs (APA 2013 pp. 87-8).

Of the characteristic features of psychosis, delusion is the feature which in this thesis I am principally concerned. That is, in each of the ‘abnormal’ (i.e. clinically significant) case studies which I examine, the occurrence of delusion has a central role. However, it is important to note that the five key features listed above ought not to be thought of as discrete or fundamentally disconnected from each other, or indeed, from the circumstances of a patient’s wider cultural or social environment and personal life. For instance, regarding the attribution of delusion and hallucination, The Diagnostic and Statistical Manual of Mental Disorders, fifth edition, i.e. DSM-5, stipulates that sensitivity is required as to the patient’s experience of cultural and social norms. For example:

An individual’s cultural and religious background must be taken into account in evaluating the possible presence of delusional disorder. The content of delusion also varies across cultural contexts. (APA 2013 p. 93.)

Likewise, delusion, hallucination, mood and related downstream behaviours are often all interwoven in the same case of psychotic disorder. And the case I introduce shortly of perceptual delusional bicephaly suffered by a patient who has a long term history of problems including the diagnosis of schizophrenia exemplifies this point. Or again, from DSM-5 re delusional disorder:
Hallucinations, if present are not prominent and are related to the delusional theme ... (APA 2013 p. 90.)

Anger and violent behaviour can occur with persecutory, jealous, and erotomanic types [of delusion]. (APA 2013 p. 92.)

Thus what can be taken away from this is that in the everyday world a clinical case of delusion is not something which occurs in a vacuum, but it is a problem which surfaces within and out of a particular human being’s life. And I suggest that this point alone supplies a sufficiently good reason for adopting in this thesis a wider more applied philosophical approach.

So within the context of delusion as a clinical, social and personal concern what precisely is meant by the term delusion? In the following section, I outline the characteristic features of delusion as these relate to the three delusions I use as case studies. Namely, erotomanic delusion, Capgras delusion, and perceptual delusional bicephaly. However, in summary, delusion can be defined as a conviction [that \( x \) is the case] which is resolutely maintained despite all argument and evidences to the contrary. Yet a problem with this is when does a conviction stop being a conviction and become a delusion. As DSM-5 observes:

The distinction between a delusion and a strongly held idea is sometimes difficult to make and in part depends on the degree of conviction with which the belief is held despite clear or reasonable contradictory evidence regarding its veracity. (APA 2013 p. 87.)

Yet then there is the problem of what counts as ‘reasonable evidence’. For instance, Smith believes with all of his heart that there is an abundance of evidence that God is here with us, whereas Jones believes with all of his heart that the evidences demonstrate that there is no such thing as God. Who is right? Which one of them is delusional?

So in addition I append the stipulation that a conviction stops being a conviction and becomes a (clinical) delusion when it starts to have a serious negative impact on a person’s life. For sure, there are awkward cases and counterexamples. For instance, Wendy is adamant that Peter lives in Neverland and she lives happily ever after
maintaining this view, whereas Ebenezer is adamant that fairytale stories are humbug and he suffers very negative consequences as a result. Again, which one of them if either is (clinically) delusional? Nevertheless, there are very many more non-controversial cases to draw attention to, and I maintain that as a rule of thumb the above stipulation is a meaningful enough generalisation to make.

The two competing theoretical models

So taking delusion as it occurs in cases of clinical psychosis, while approaching it as a philosophical as opposed to a purely medical concern, the prevailing view is that there are two competing models for explaining why delusions come about.

On the one hand, there is the empiricist model, which states that, in any given case, a delusion is explained by the prior occurrence of an (clinically) abnormal experience, and this view is further divided into two competing groups. The empiricist one-factor approach maintains that the fact of the abnormal experience is both necessary and sufficient to account for any case of delusion, that is abnormal experience is always the sole significant explanatory cause (e.g. Maher 1974, 1999, 2005). Alternatively, the empiricist two-factor approach maintains that although in all cases of delusion abnormal experience is a necessary explanatory factor, it is not sufficient, and that subsequent to the abnormal experience abnormal cognitive processing is the jointly necessary explanatory cause (e.g. Coltheart et al. 2010; Coltheart et al. 2011).

On the other hand, there is the rationalist model of delusion which states that, in any given case, a delusion is explained in terms of abnormal cognitive processing alone. The foremost proponent of the view is John Campbell, whose rationalist approach to delusion (e.g. Campbell 2001) demands that never in any case of delusion can (clinically) abnormal experience – qua hallucination – have an explanatory role.

I say more about the three prevailing approaches to delusion shortly, then discuss in some depth their respective strengths and limitations in chapter two. However, the point to take away for now is that, re the conventional way of thinking, the problem of explaining delusions is fundamentally binary, in so much as if either empiricism or rationalism is right then the other approach must be wrong. However, as I argue in chapter two, each of the prevailing theories has strengths but also weaknesses, in that one approach seems to deal with one sort of delusion but not another, whereas the other approach seems able to meet the latter sort of case but not the former.
Yet, for this reason at least, it seems not unreasonable to query whether the binary approach is not in fact ideal. For instance, maybe there is logical space for a third way of thinking about the problem of delusion which effectively slots between the conventional empiricist and conventional rationalist approaches. This is pretty much the position occupied by the Dennettian and Prediction Error (D-PE) account, the two sides of which I outline later in this chapter, before developing the combined model in chapter three, and which technically, effectively by default, tends towards a rationalist account of delusion, but which in some ways shares significant ground with an empiricist approach.

So to sum up this introductory section, it can be said that the philosophical problem of delusion as it occurs in a clinical setting can be described as having two sides. First, there is the problem of identifying the theory of delusion which most comprehensively addresses the fact of delusion and its associated features. And second, on the basis of this examination, there is the question of specifying which, if either, of the prevailing explanatory models of delusion works the best.

**Features of Delusion: Three Case Studies**

A diagnosis of delusion is always a specific case affecting a particular person, hence the subjects in each of the case studies I examine is granted a name. Nevertheless, in general, delusions are talked about in terms of their having (or not) various characteristic properties, but even this remains dependent on assuming a particular point of view. For instance, a neuroscientist investigating a delusion may be interested in fine-grained patterns of biological sorts which present as very different to the softer, wider, behavioural patterns which are important to the approach taken here. This is not to suggest that one type of pattern is right or more right than the other, it is just that each exhibits meaning from a different explanatory perspective. So given this caveat, the features of delusion which are more relevant to me here are as follows.

First, and most generally, the definition of delusion as it appears in DSM-5 begins by observing that: Delusions are fixed beliefs which are not amenable to change in light of conflicting evidence (APA 2013 p. 87). The term ‘belief’ is being used here in the standard sense of a propositional state, i.e. a mental attitude state the content of which can be expressed via a sentence of natural language, and which is typically glossed
as a belief that $p$, e.g. Jill believes that it is going to rain. Yet this is not the only way in which the word ‘belief’ can be used. And, as I outline later in this chapter, Dennettian theory uses ‘belief’ to specify something different to a proposition. So when talking about delusions, I use the word conviction rather than belief.

Second, it is recognised that delusions tend to correspond to certain characteristic types, and DSM-5 lists some of the more prevalent types of delusion with which patients present (e.g. APA 2013 pp. 90-1). One delusion specified by DSM-5 is that of an erotomanic type, i.e. erotomanic delusion (ibid.), which is a delusion that I use as a test case, along with Capgras delusion and perceptual delusional bicephaly. I introduce each of these case studies in a moment, but, in brief, erotomanic delusion has as its central theme a subject’s (false) conviction that another person is in love with her (APA 2013 pp. 87, 90); Capgras delusion has as its central theme a subject being convinced that a person known to him has been supplanted by a (tantamount) identical replacement (e.g. Alexander et al. 1979); and perceptual delusional bicephaly has as its central theme the subject’s conviction that he has, i.e. that he sees, hears, and feels an (ostensibly) autonomous second head (Ames 1984).

Third, as even this brief introduction to the case studies illustrates, although delusions are essentially a cognitive phenomenon, some cases of delusion include an experiential factor. Accordingly, another way of categorising delusions is by the experiential content they have. As regards this condition delusions can be divided into three groups. Zero (experience) delusions are delusions which (characteristically) have no significant experiential element, e.g. erotomanic delusion. Weak delusions are delusions which (characteristically) have a mildly anomalous experiential element, such as a nagging salience or a feeling that something is not quite right (cf. Kapur 2003). And (very) strong delusions are delusions which (characteristically) feature a significant hallucinatory element, which as the case of perceptual delusional bicephaly demonstrates can be a very significant anomaly indeed.

Fourth, delusions can be categorised by the plausibility (or implausibility) of their contents (APA 2013 pp. 90-1). That is to say, some delusions are not impossible, but are nevertheless implausible, for instance, an erotomanic conviction that a notable celebrity is in love with one might not be unthinkable, but in certain circumstances will be manifestly improbable all the same. Then again, other delusions are not impossible as such, but they are bizarre, for instance, a case of Capgras delusion in which the patient maintains that his family has been replaced with tantamount identical replicas is not
entirely impossible, but is fantastical. Then there are delusions which are flatly impossible, for instance the conviction that one has acquired a fully functioning second head.

Fifth, delusions can be categorised by how they fit into a patient’s wider mental landscape. For instance, according to whether the content of the delusion is monothematic or polythematic. A delusion is said to be monothematic when it is effectively limited to either a single conviction, or to a narrow set of convictions around a single theme, and it may be that the delusion is the only significant mental issue with which the patient presents, i.e. that he or she is an otherwise ‘normal’ (healthy enough) person. It is a pattern which is typically associated with diagnoses of delusional disorder (APA 2013 pp. 90-3), and with monothematic delusional syndromes – where ‘syndrome’ basically means a pattern of symptoms which is recognised to occur as regards different subjects and which has a particular delusion as its central theme. For instance, a monothematic delusional disorder featuring delusions of an erotomanic type is sometimes referred to as ‘erotomania’ or ‘de Clerambault’s syndrome’ (after Gaëtan Gatian de Clérambault, who first recorded it), whereas a monothematic delusional disorder involving Capgras delusion is sometimes referred to as Capgras syndrome (after Joseph Capgras). Alternatively, a delusion is said to be polythematic if its contents are wide-ranging and complex. This is a pattern of disorder more often found in cases of schizophrenia, in which a patient may present with an array of mental and physical concerns (APA 2013 pp. 99-102). However, a delusion which centres on a single specific theme, i.e. which is effectively a monothematic delusion, can occur in conjunction with schizophrenia. The subject of the case study involving perceptual delusional bicephaly is a case in point (Ames 1984).

Sixth, according to different theories of delusion, delusions are variously said to co-occur with certain other factors, which (factors) are variously held to play (or not) a significant explanatory role. There are three such factors which are relevant to this thesis. They are the root cause of the delusion; the factor of rationality (or irrationality); and experience.

Specifically, the root cause of delusion stipulates whether a delusion is the consequence of an underlying organic (biological, specifically neuropathological) disorder or whether it has a psychodynamic (personal, motivational, emotional) cause (Ellis and de Pauw 1994). As regards this point, each of the prevailing theoretical approaches adopts a rigid stance, in that they all insist that in any case of delusion organic disorder is the necessary and sufficient underlying explanatory cause of the
delusion (e.g. Maher 1974; Coltheart et al. 2011; Campbell 2001). In contrast, a D-PE view offers a more flexible proposal in that it awards neuropathology a contingent role as a significant explanatory cause (see chapter four).

As regards rationality, the empiricist two-factor account and Campbell’s rationalist account both hold that in all cases of delusion deep irrationality is by necessity a key explanatory factor (e.g. Coltheart et al. 2011; Campbell 2001), while the empiricist one-factor account and the D-PE account both maintain that deep irrationality is not a factor (e.g. Maher 1974, 1999, 2005; see chapters three and four), that is to say, they insist that in any typical case a subject suffering from delusion is a fundamentally rational human being.

And with regard to experience, (as previously mentioned) each of the prevailing theoretical accounts assumes an uncompromising stance re the role played (or not) by anomalous experience in any case of delusion, whereas a D-PE approach takes abnormal experience to be a contingent explanatory cause.

Seventh, delusions can be classified concerning the way in which they change over time. The rule here is that there is no single rule. Some cases of delusion have a very short duration, e.g. occurring only in conjunction with an episode of migraine (e.g. Ellis and Pauw 1994). Other cases remain fixed for many years (e.g. Alexander et al. 1979). And some cases continue for a period of time, but during that time the delusion changes, e.g. the focus of an erotomanic delusion may (repeatedly) shift to a different object (e.g. Menzies et al. 1995), or maybe what starts out as a delusion with a specific content becomes increasingly complex, which may coincide with a progressive deterioration in the patient’s overall mental state (e.g. Lipson and Mills 1998).

Lastly, when grouped together, this list of characteristics of delusion adds up to a data set of features of delusion. And the three cases of delusion which I use to discuss the theories of delusion have been chosen because together they represent a broad cross-section of the data set. Thus if any theory of delusion is to comprehensively address the idea of (clinical) delusion, then it at least needs to explain each of the cases of delusion outlined below.

Case study 1: Erotomanic delusion

In erotomanic delusion, a person (patient, subject) holds the persistent false belief that a second person (the object or victim) is in love with them and engaging in intimate communication with them and/or that they and the object are destined to be together if
only they were to meet (Davey 2008). Hallucinations rarely feature. However, should hallucinations figure, they tend to be a later development, arising as the disorder is gaining in complexity, and they typically involve dreamy somatic-type experiences, perhaps in situations around sleep (e.g. Kennedy et al. 2002).

As regards the profile of a patient who presents with erotomanic delusion, there is overwhelming agreement between studies. The subject is typically female, although in cases involving violent behaviours, the subject is almost always male. And subjects are variously described as being shy, lonely, (sexually) inhibited, withdrawn, isolated individuals; having poor levels of emotional, social and occupational support and functioning; and of unattractive physical appearance (e.g. Menzies et al. 1995; Lloyd-Goldstein 1998; Meyers 1998; Fitzgerald and Seeman 2002; Kennedy et al. 2002). In contrast, the object of the delusion is characteristically (but not necessarily) a person of higher social standing than the subject, for instance a clinician or a celebrity (e.g. APA 2013 p. 91; Fitzgerald and Seeman 2002).

Concerning the clinical classification of erotomanic delusion, it is generally accepted that it can occur as a primary disorder (i.e. happening ‘out of the blue’, in the absence of a prior disorder or pathology), or a secondary disorder (i.e. in connection with a prevailing clinical concern). Furthermore, studies confirm that there are noticeable behavioural differences between primary and secondary cases. For example, primary erotomanics are more likely than secondary to have delusions about single celebrity objects; view the ‘relationship’ in terms of a pure, idealised, ‘spiritualised’ connection; be less likely to have delusions with sexual content or engage in affective or forensic behaviours (Menzies et al. 1995; Lloyd-Goldstein 1998; Kennedy et al. 2002). On the other hand, erotomanic delusion occurring within the context of a wider psychiatric disorder such as schizophrenia is more likely to pertain to multiple objects, have sexual content, and result in forensic behaviours perpetrated against the object of the delusion or their family (Menzies et al. 1995). Nevertheless, this ought not be taken to mean that every single secondary erotomanic subject is dangerous or that antisocial, dangerous and forensic behaviours are never perpetrated by primary erotomanics. For instance in the so-called ‘Tarasoff case’, Prosenjit Poddar, who at the outset was a seemingly normal-enough overseas student, developed a monothematic erotomanic-type fixation on a fellow student, Tatiana Tarasoff, who he subsequently stabbed to death (Lipson and Mills 1998).
As regards the root cause of erotomanic delusion, as things stand, there is no clear evidence to suggest that there is any significant correlation between (primary) erotomanic delusion and neuropathology, i.e. a significant organic change in the brain, which is not to suggest that erotomanic delusion is never reported in cases involving a prior nonspecific brain disorder, such as dementia (Brune 2001). A possible counterexample to this might be a case in which a male subject was admitted into a secure psychiatric unit on account of erotomanic type and related stalking behaviours, whereupon he was found to have a large and previously undiagnosed lesion of his brain, and during the treatment of which lesion a cessation of the erotomanic type symptoms took place (Farnham et al. 1997). However, in this case, if a correlation exists between the brain lesion and the delusion, then the nature of that correlation is highly unclear, especially given the wider special and difficult circumstances which the patient at the same time faced.

On the other hand, it is widely considered that there is a meaningful connection between instances of erotomanic delusion and various special or difficult psychological, social, and especially interpersonal concerns with which the subject contends (Meyers 1998). For instance, erotomanic delusions are variously said to be: ‘a variant of pathological mourning’ (Evens et al. in Meyers 1998); ‘part of a paranoid disorder’ (Feder 1973 and Goldstein 1987 in Meyers 1998); ‘a reflection of poor reality testing’ (Meloy 1989 and 1997 in Meyers 1998); related to ‘morbid jealousy’ (Berry and Hayden 1989 and Munro et al. in Kennedy 2002); a compensation for ‘unmet narcissistic needs’ (Lipson and Mills 1998; Enoch 1991 and Kraeplin 1921 in Kennedy et al. 2002). For instance:

Erotomanic delusions can involve denied self-love projected onto a superior other, or may be a reaction formation against social and sexual isolation or childhood traumas, as Kraeplin once suggested “a psychological compensation for the disappointments in life” (Kennedy et al. 2002 p. 4).

Finally, regarding the development and duration, i.e. clinical course, of the disorder, erotomanic delusion presents as something of a mixed bag (e.g. Kennedy et al. 2002; Fitzgerald and Seeman 2002). For instance, the delusion can have a relatively short duration (e.g. Farnham et al. 1997); it can stabilise as mood stabilises when antipsychotics are begun (e.g. Fitzgerald and Seeman 2002); it can endure, unchanging
for many years (Menzies et al. 1995); it can deteriorate fairly rapidly into compounding mental disorder, and possibly forensic behaviours (Menzies et al. 1995), and the above mentioned Tarasoff case would be an example of the latter situation (Lipson and Mills 1998).

So given the characteristics associated with typical cases of erotomonic delusion as they are outlined here, erotomonic delusion can be classified as a case involving zero significant experiential factor and for which the evidence points towards an explanation in terms of a psychodynamic approach. Accordingly, in chapter four, I apply these ideas to the case study of Patsy, who is a fictional subject, but whose story tracks a developing monothematic erotomonic type delusional disorder and serves to exemplify many of the features of the delusion presented above.

**Case Study 2: Capgras delusion**

Capgras delusion is not an especially common disorder (e.g. Bentall 2003 [2004] p. 299). All the same, it is a delusion which receives a great deal of attention in the academic literature (e.g. Alexander et al. 1979; Ellis and Young 1990; Ellis and de Pauw 1994; Maher 1999; Campbell 2001; Hohwy 2004; Coltheart et al. 2010; Coltheart et al. 2011).

The key characteristic of Capgras delusion is that the subject is convinced that a person known to them has been replaced by a duplicate. Often there is a close emotional tie between the subject and the object of the delusion, for instance, the object may be a spouse or child (e.g. Alexander et al. 1979). However, this is not necessarily the case, and the object may be quite tangentially related to the subject, for instance, a neighbour, an associate, or an attending nurse (e.g. Ellis and de Pauw 1994).

As regards the role of experience in Capgras delusion, a scan through the literature reveals that there is general agreement that patients suffering with Capgras delusion characteristically present with atypical feelings of reduced or absent affect as regards the object of their delusions (e.g. Maher 1999; Ellis and Young 1990; Coltheart et al. 2010), which serves to classify Capgras delusion as a case of delusion in which weak anomalous experiences characteristically occur. However, what is not particularly clear is whether the experiences in question are a key upstream explanatory feature of the delusion; an associated feature which perhaps shares a causal root with the delusion; a consequence of the delusion; and whether they (continue to) play (or not) a significant downstream role in any given case. From a purely diagnostic point of view the difference between these possibilities is not particularly significant. That is to say, if
one’s objective is to recognise the occurrence of a possible disorder, and if anomalous experience reliably occurs within the pattern of changes which indicates a diagnosis of Capgras delusion is warranted, then the mere observation of the occurrence of anomalous experience is sufficient to support one’s needs. However, from a theoretical standpoint of explaining or categorising Capgras delusion, the relationship of experience to the occurrence of the delusion is important. At the very least, an empiricist approach ought to demand that the anomalous experience is a prior causal factor in explaining the delusion (e.g. Maher 1999; Coltheart et al. 2010), whereas John Campbell’s rationalist account pushes the anomalous experience into a less crucial downstream role (Campbell 2001).

Concerning the circumstances in which Capgras delusion happens, this presents as something of a mixed bag. For instance, Capgras delusion is recognised to present as a monothematic delusional disorder (i.e. Capgras syndrome), in which the delusion is the sole or central concern (Alexander et al. 1979). Alternatively, the Capgras delusion may be but one feature within a complex psychiatric disorder (e.g. schizophrenia) (e.g. Ellis and de Pauw 1994). In addition, Capgras delusion has been observed to occur in conjunction with certain medical problems (e.g. delirium) (e.g. Ellis and de Pauw 1994) and/or various structural deficits in the brain (e.g. traumatic damage) (Alexander et al. 1979). As for its duration, episodes of Capgras delusion are known to vary between those which last for a short spell, e.g. 48 hrs in conjunction with an attack of migraine (e.g. Ellis and Pauw 1994), and those which persist for a much longer period of time, e.g. many years in conjunction with a lasting head injury (e.g. Alexander et al. 1979).

As to the underlying explanations of Capgras delusion, there are a number of competing theories in play. For instance, Capgras delusion is variously said to be the consequence of problems involving: misidentification (Alexander et al. 1979; Ellis and de Pauw 1994), especially as regards the visual modality (Bentall 2003 [2004] p. 299), since there are reports of a subject speaking with the object via telephone with no corresponding conviction that the object is anyone else but whom they actually are (p. 310); motivation, with especial appeal made to the subject having diminished autonomic responses (e.g. Ellis and Young 1990; Coltheart et al. 2010; Coltheart et al. 2011); memory (Alexander et al. 1979; Campbell 2001); existing psychiatric illness (Bentall 2003 [2004] p. 311).

With regard to the aetiology of Capgras delusion, by and large the prevailing views divide into two camps, namely, those endorsing a psychodynamic explanation and
those endorsing an organic account (Ellis and de Pauw 1994). The more traditional view is that Capgras delusion is a psychodynamic disorder (ibid.). However, more recently, it has been proposed that Capgras is at root an organic disorder (ibid.). And while the fundamental organic deficit has not yet been clearly specified, the options mooted are basically that the root of the delusion is to be found in either a ‘hard neuropathology’, i.e. structural damage or deficiency in a region of the brain, and about one third of Capgras patients are observed to have abnormalities of this sort (Bentall 2003 [2004] p. 310), or a ‘soft neuropathology’, i.e. a physiological or functional problem, such as a neurochemical disparity (Ellis and de Pauw 1994), but the number of subjects affected in this latter way is hard to determine, not least because the pertinent functions can be difficult to assess and tend to be affected by many different factors, such as medications and stress.

However, even if an organic anomaly is detected in the brain of a subject presenting with Capgras delusion, as was the case with the role of experience, from a theoretical perspective, the nature of the relationship between the anomaly and the delusion – and specifically the idea that said relation is explanatorily significant – cannot be taken for granted, accordingly any proposed relationship needs to be mapped before conclusions can be set. Plus it cannot be disregarded that there remains a non-trivial number of cases in which no significant organic abnormality is detected—and perhaps given the circumstances of a particular case an assumption of such an abnormality would not be warranted. And so while fully acknowledging that there are clear limits to what the lack of an observed abnormality can actually establish (Ellis and de Pauw 1994), the fact remains that respecting the data which we have to work with at present, there is a set of cases in which no abnormality features and so which needs to be explained, or at least explained away. Therefore, as Ellis and de Pauw observe, even if one favours, as they do, the idea of an organic basis to Capgras delusion, psychological explanations for the disorder ‘should not be dismissed out of hand’ (p. 320).

So with this in mind, I use two case studies to illustrate Capgras delusion. Case study 2a leans towards an explanation in terms of psychological factors. Case study 2b is a case in which serious traumatic injury to a subject’s brain seems to bear a relationship to the Capgras delusion which subsequently develops.
Case study 2a
This is the original case described by Joseph Capgras and J. Reboul-Lachaux in 1923, as it is outlined in English by Richard Bentall (Bentall 2003 [2004] pp. 309-310). It involves a subject, Mme M, who, following the death of her twin sons and the breakdown of her marriage, undergoes a series of delusional misidentifications. Specifically, at the graveside, observing her child’s small coffin being lowered into the ground, Mme M becomes convinced that the child in the coffin is not in fact her child; then that her husband has been replaced by an imposter, which imposter is then repeatedly replaced by yet further imposters; then the grandiose conviction that she is a member of the French aristocracy, but has herself been replaced by an imposter; and then that the WW1 tunnels beneath Paris are filled with people, all of whom have been replaced by imposters, and that the aircraft flying over Paris do so with the objective of driving more people below so that these people too can be replaced in due course. As Bentall observes, it is ‘not hard to see why Capgras offered a psychoanalytical account of Mme M’s illness’ (p. 310).

Case study 2b
According to its authors (Alexander et al. 1979), this is a case study which has had ‘considerable influence’ in the ongoing conversation re Capgras delusion (Alexander and Stuss 1998). The subject is a man, to whom I give the name Mr A. And concerning the events recorded in the case history, Mr A is 44 years old, married, with children, and prior to a road traffic accident in which he sustained severe head injuries, Mr A was a normal healthy adult, with a good family relationship, albeit with some minor history of mild delusional ideation, but nothing that would place him on the clinical map. However, as a result of the accident, Mr A sustained traumatic damage to his brain and functional impairments relating to motor skills, insomnia, motivation, memory and misidentification, including that of his immediate family, who Mr A became convinced had been replaced by duplicates. Gradually, Mr A’s condition improved. After three years he exhibited no deficits in recent memory and could recognise friends, but he still held the conviction that his ‘original’ family had been replaced with a wife of almost identical appearance, and the same number of children with the same genders, names, appearances etc. — although he believed that the children in his original family were about one year younger than those of his replacement family. Nevertheless, Mr A maintained a positive attitude towards ‘both’ families, although he made no attempt to
contact the original one. Moreover, although Mr A exhibited no obvious deficit of logical competences, he could not use the logic of situation to amend his beliefs. That is, Mr A acknowledged the implausibility of his story, but nevertheless insisted that it was correct, and when asked probing questions his responses became vague and evasive: ‘You’re asking difficult questions’ (Alexander et al. 1979 p. 335). Months later the subject’s condition remained unchanged. It was, however, proposed that his continuing problems may have something to do with his diminished motivational drives: ‘part of the difficulty appeared to stem from pathologic unconcern and lack of motivation to rectify the situation’ (p. 335). Finally, wider-ranging mental or emotional symptoms were never a factor: ‘at no time after the injury were suspiciousness, paranoia, hallucinations, thought disorder, or ideas of reference noted, and he never displayed agitation or anger to anyone, including his wife’ (p. 334). Thus here is a case in which the occurrence of an organic deficit, i.e. actual traumatic damages to brain tissue and associated functional problems, correlates with the onset of Capgras delusion in a patient who, with the exception of post-trauma motivational issues, otherwise seems (eventually to recover) to be normal enough.

**Case Study 3: Perceptual delusional bicephaly**

While delusions of bicephaly which feature somatic-type hallucinations of a second head are rare but not unheard of, the case of delusion described in this case study is, according to its author, tantamount unique (Ames 1984). The uniqueness of the case rests with the fact that the delusion involves not merely somatic hallucinations of a second head, but also a complex of very strong perception-like visual and auditory hallucinations, which lead to the second-head presenting as a mind-independent entity with a distinctive personality of its own with which it engages in hostile and competitive exchanges with the subject, and on account of which hostilities the subject eventually tries to ‘eliminate’ the head by shooting it dead, in the process seriously injuring his actual head, and which injuries eventually contribute to his death two years later on.

As outlined above, it is a characteristic feature of some cases of delusion that they have a hallucinatory factor which is clearly related to the theme of the delusion. And while the hallucinations which occur in different clinical situations are recognised to have varying characteristics as regards, modality (e.g. visual, auditory etc.); intricacy (e.g. from simple tickles, lights or buzzes to complex epic scenes); spontaneity (e.g. they are involuntary or they are something over which the subject has some control);
strength (e.g. they are perceptually weak, or vividly strong); and their seeming to occupy a loci in either the mind or the world (e.g. Knight et al. 2008; ffytche 2013). In the context of psychosis, hallucinations are pointedly defined as being ‘perception-like experiences that occur without an external stimulus. They are vivid and clear, with the full force and impact of normal perceptions, and not under voluntary control’ (APA 2013 p. 87). And indeed, although the specific and dramatic contents of Ames’ case study make it a memorable example, the fact is that hallucinations which are complex, vivid, disagreeable, persecutory and bizarre are not unusual in cases of schizophrenia (APA 2013 pp. 87-8, 99-105; Knight et al. 2008), which is the psychiatric disorder with which the subject of the case study contends (Ames 1984). Therefore, it is reasonable to use Ames’ case study as a test case for delusions which involve a hallucinatory factor of a (very) strong kind.

So turning to the case study, it relates to a man, to whom I give the name Mr P. At the time of admission, Mr P is 39 years old. He has a six year clinical history of psychiatric problems, some of which have required in-patient care. Two years prior Mr P was the driver of a car which was involved in an accident in which his wife died. And on admission, Mr P presents with a self-inflicted gunshot wound involving significant traumatic injury to his head. Mr P is subsequently diagnosed to be suffering from schizophrenia (Ames 1984).

The circumstances around and leading up to Mr P’s admission centre around the theme of his late wife and involve a raft of hallucinations, the nub of which involve Mr P’s deeply held conviction that he has and is in communication with a deeply unpleasant second head (ibid.). Mr P describes being able to see, feel and hear the second head on his shoulder. He also claims to have heard the voices of Jesus and Abraham conversing, and confirming that he does indeed have a second head. Mr P is convinced that the head is that of his late wife’s gynaecologist, with whom he believes his wife had been having an affair. And Mr P describes the head as relentlessly trying to ‘dominate’ his own head, referring to itself as the ‘kingpin’ and claiming that it is going to take Mr P’s wife away. Mr P recalls that, after about three weeks of this, he had had enough and that he considered taking an axe to the head before deciding to shoot it off – which act results in the head dangling by a thread, at which point Mr P claims to have been feeling good, being no longer able to feel the second head, before he passed out (ibid).

Following admission, over a period of three to four weeks, and subsequent to relevant surgical and medical interventions, Mr P’s willingness to discuss the head
waned until he would no longer talk about it. After eight weeks, Mr P claimed that his symptoms had disappeared, and he was discharged. However, Mr P continued to have problems relating to alcohol intake, and about a year later he stopped taking his prescribed antipsychotic medications and was readmitted, whereupon he reflected on his prior admission:

Q. Could you see the other head?
A. Yes.
Q. You felt it, or you could see it?
A. I could see it.
Q. And the voices were coming from the other head?
A. From that head and my own head too.
Q. Whose voice was it?
A. It was the voice of my wife’s doctor.
Q. What was he saying to you?
A. He had an affair with my wife and he reckoned he was going to take her off me and all that kind of talk and I got antagonistic towards him and I decided to do something about it and I shot myself.
Q. Since that episode have you had anything like that?
A. Not really. I have not heard any voices for about two years, ever since I shot myself and I haven’t had any ideas that I have two [heads] again. When I shot myself it fixed it up.

(Ames 1984 pp. 193-4)

About a year later, Mr P is readmitted in a comatose state and two-days later he dies. Whereupon, post-mortem examination reveals significant and widespread damages to his brain on account of the old gunshot injury and an ongoing infection which is likely related to the original wound (p. 194).

This concludes the four case studies of delusion which I use to discuss the accounts of delusion in the following chapters. To recap, they are:

The case of Patsy – erotomanic delusion.
The case of Mme M – Capgras delusion, a psychological explanation.
The case of Mr A – Capgras delusion, involving an organic pathology.
The case of Mr P – perceptual delusional bicephaly.

The Prevailing Approaches to Explaining Delusion

There are two established theoretical approaches to the problem of delusion. These are the empiricist and the rationalist explanatory models. The empiricist model is essentially the proposal that in any (i.e. by necessity in every) case of delusion anomalous experience is the key explanatory factor. The rationalist thesis is essentially that in any case of delusion abnormal cognitive processing is the sole significant explanatory cause. The two models are standardly applied via three prevailing theories of delusion. The theories are: the empiricist one-factor account (e.g. Maher 1974, 1999, 2005); the empiricist two-factor account (e.g. Coltheart et al. 2010; Coltheart et al. 2011); and John Campbell’s rationalist account (Campbell 2001).

The three prevailing theories can be summarised as follows. As regards both empiricist accounts the fundamental principle is that a subject experiences something which is significantly odd, which he or she struggles to understand, and it is in the process of these cognitive efforts that a related delusion develops. The difference between the empiricist approaches turns on whether or not the cognitive struggles are themselves deemed to be clinically abnormal, i.e. the anomalous consequence of a deep underlying pathology. The one-factor theorist says, No, cognitive processing is normal (e.g. Maher 1974). The two-factor theorist says, Yes, abnormal cognitive processing is a jointly necessary explanatory cause (e.g. Coltheart et al. 2011). On the other hand, Campbell’s rationalist account maintains that anomalous experience is never a factor and that a deep cognitive abnormality effects downstream changes which lead to the case of delusion developing (Campbell 2001).

Thus, sketched in this way, the claims submitted by the three prevailing theories seem to present as a straightforward disagreement. The empiricist says Yes to experience whereas the rationalist says No. In conjunction with the rationalist and two-factor empiricist endorsing the idea of deep irrationality, and the one-factor empiricist denying deep irrationality a role. However, caution is required. For despite the seeming straightforwardness of the theories’ respective positions, the prevailing empiricist and rationalist approaches to explaining delusion actually rest on some very different assumptions. Specifically, they draw on two rival accounts of experience. So not only is
comparison between their central claims not as straightforward as it seems, since it is not just a matter of comparing like with like, but the different starting points they have regarding experience give rise to different and theoretically significant consequences regarding what each of the approaches can and cannot explain.

The empiricist approach

If we start by considering the empiricist’s case. Essentially the empiricist model rests on one version or another of the causal theory of experience (e.g. Grice 1961; Ayer 1976; Strawson 1979; and many others). The nub of the idea is that mental phenomena, such as experience (consciousness) and belief (cognition), are the result of causal changes in the brain. And consciousness, whether veridical perception or perceptual hallucination, is the actual consequence of one kind of change, i.e. perception and hallucination are common in kind (e.g. Macpherson 2013; ffytche 2013); whereas the changes which realise cognition are fundamentally different, which is to say consciousness and cognition are each founded on a discrete set of neural correlates (e.g. Dennett 1991 [1993]; Cohen and Dennett 2011). Furthermore, concerning these phenomena, it is held that consciousness is primary and that the contents of consciousness supply the contents of cognition. That is to say, one’s beliefs are formed on the basis of one’s experiences. However, some theorists propose the relationship between experience and belief isn’t wholly one way, and that cognitive content can or does affect or penetrate the content or character of consciousness (e.g. Siegel 2010, 2012; Macpherson 2012). Yet even if it were the case that cognition can and does penetrate consciousness, it would still be the case that within this model consciousness and cognition are different kinds.

So thinking about delusion, the empiricist one-factor thesis is the view that in any given case of delusion a clinically anomalous experience, e.g. a hallucination, occurs, and it is in trying to explain this anomaly via fundamentally normal cognitive processes that the false conclusions characteristic of the different types of delusion come about (Maher 1974, 1999, 2005). In contrast, the empiricist two-factor thesis maintains that in any given case of delusion a clinically anomalous experience is the first significant explanatory factor and about which experience there is then a struggle to understand via cognitive processing which is itself clinically abnormal, i.e. where the cognitive failure is held to be rooted deep inside the brain, and where the skewed effects of this abnormal processing result in the false conclusions characteristic of different delusions (Coltheart
et al. 2010; Coltheart et al. 2011). Thus there are two competing empiricist theories of delusion, which disagree about the role awarded to (anomalous) cognition, but which agree on the role of (anomalous) experience. Moreover, given their shared empiricist background, they are not only in agreement about what experience does, but are also in broad agreement about what experience is. That is, experience is actual phenomenal consciousness which whether in perception or hallucination is composed of the same sort of stuff, i.e. is common-in-kind. Accordingly, if we use the case of Mr P as an illustration, if Mr P were to perceive (veridically) a second head, and if Mr P were to have a (very) strong hallucination of a second head, then in both instances what-it-is-like for him to have the experience would be identical. That is, although subsequent reality testing should in theory be able to separate the two states, there would be nothing about the experiences per se which could tell them apart.

Should it be queried, does this mean that for an empiricist a hallucination has to be identical to perception? The answer would be no and yes. On the one hand no, because, as has already been noted, hallucinations occur in different stripes, some can be weak and dreamy and so dissimilar to normal perception, whereas others can be strong, multimodal and (very) perception-like. Nevertheless, it has also been noted that as regards psychosis, the type of hallucinations recognised as having clinical significance are those of a (very) strong kind. So at least in the case of Mr P, the hallucinations at point need to be approached as being (very) perception-like in kind. On the other hand, the answer is yes because regardless of whether a hallucination is weak and dreamy or strong and (very) perception-like, within this framework hallucination is fundamentally composed of the same kind of stuff as any other experience, i.e. the stuff of actual phenomenal consciousness, in virtue of which the subject experiences the features of his or her world.

And should it be queried, must the empiricist about delusion defer to a causal view of experience? Again the answer is no and yes. No because technically speaking there is no necessary reason why an empiricist approach to delusion must rest on a causal construal of experience. There are goodness knows how many competing theories of experience in circulation and in principle the empiricist could favour any one. However, in order to maintain an empiricist position about delusion there are certain stipulations which need to be met. Namely, within the theory experience would need to be construed as being the actual product of empirical changes in the living brain (though perhaps this might not be completely necessary so long as the theory
upholds the following conditions); have a fundamental nature which is dissimilar to that of cognition (though perhaps not entirely closed off from cognition); and happen prior to cognition in the explanatory chain (hence work as the key reason why one arrives at the cognitive contents one does). However, the answer to the query is also yes because, while there are alternative ways of thinking about experience which are not (strictly) causal in nature, the problem for the empiricist about delusion is how well (if at all) they concord with principles of the empiricist view. For instance, a relationist approach to experience does not appeal to the idea of a chain of causality between the observed and the observation, but to a direct association of the two (e.g. Snowdon 1980-1, 1990-1, 2008; Haddock and Maepherson 2008 [2011]). However, relationism about experience is the theory on which John Campbell’s rationalist account of delusion rests so not the empiricist thesis. Likewise, Jakob Hohwy (Hohwy 2013) and Daniel Dennett (Dennett 1991 [1993]) both propose theories of experience in which causality features, but in a much looser way than is conventionally construed. Yet while Hohwy’s approach awards a role to actual experience, it is not the case that experience is essentially prior to belief (Hohwy 2013 pp. 25-6, 117-38). While Dennett’s view explicitly rejects the idea of actual experience, and so categorically experience is not prior to belief (e.g. Dennett 1991 [1993] pp. 321-68, 369-411, 431-55). Thus while there certainly are alternatives to the causal approach to experience, any alternative would need to be clearly outlined and independently examined to determine not only how well it addresses the problem of delusion, but how well it does so inside the empiricist about delusion’s school of thought. That’s why until then the answer to the above query is yes, one version or another of the causal view of experience can be regarded as the hypothesis supporting the empiricist accounts.

Campbell’s rationalist approach

Moving on to look at the rationalist view of experience. This is not as straightforward as it was from the empiricist perspective, because at least within this thesis there is not a single rationalist school of thought. Nevertheless, as we are currently thinking about the prevailing theories of delusion, so we need to consider John Campbell’s (relationist) rationalist approach.

Campbell’s rationalist account of delusion employs the idea that the nature of human cognition is such that beneath or behind a subject’s cognitive activities there exists a core set of \textit{a priori} beliefs which work to found the basis of that subject’s
account of reality and so support (or not) every subsequent thought, inference, conclusion which the subject entertains (Campbell 2001, pp. 96-7). Campbell draws a connection between these core beliefs and Wittgenstein’s concept of framework propositions, which are said to have a different epistemic status to ordinary factual beliefs, in that factual beliefs are evidence based, i.e. their contents are consequent to prior experience and reasoning (ibid.). For instance, you see water falling from the sky, and reason to the belief that it’s raining outside. In contrast, framework propositions are the starting point of all reasoning, and so establish the environment in which any particular argument survives (or dies) (ibid.). For example, one needs to have in place certain foundational concepts about space, movement, and particular substances before one can formulate and assess the truth or falsity of propositions like water falls from the sky.

So as regards delusion, Campbell’s proposal is that in any clinical case of delusion there must be an original anomaly deep in the subject’s cognitive framework concerning one of the subject’s core beliefs (Campbell 2001 pp. 96-8). This corrupted belief is the delusion (p. 96-7). And consistent with the functional role of foundational propositions, the corrupted belief progressively works to influence subsequent cognitive processing, and as a result perhaps effects related downstream (physiological, behavioural, and/or experiential) changes, which in turn could have yet further consequences (pp. 97-9), so that in time a pattern of symptoms indicative of a particular psychosis manifests in the subject’s observable behaviours whereupon a diagnosis may be made.

So taking Capgras delusion re a spouse as an example, if within the fabric of his cognitive system a subject has a corrupted foundational proposition concerning the fact of his wife, then whenever his wife is in his perceptual environment, instead of responding in a typical manner his system might respond abnormally by triggering a relatively depressed affective response, which in turn might result in odd changes, feelings, sensations which work to supply the evidences which found the subject’s factual beliefs. At which point, the subject may advance the idea that this is not in fact my wife, it is an imposter, which establishes the pattern with which he may eventually present in a clinical setting.

Thus Campbell’s rationalist theory of delusion can be summed up by saying that in any case of delusion an anomaly in the subject’s framework propositions is the sole significant explanatory cause. Campbell argues that this does not exclude the idea that in
certain cases of delusion pertinent anomalous experiences could perhaps occur (Campbell 2001 pp. 98-9). For instance, in the case of Capgras delusion outlined above, the subject might feel a lack of emotion towards his wife, and this may disturb him, or perhaps the absence of affect could penetrate his visual experiences, so that when he sees his wife something about the experience seems different to him. Nevertheless, in line with Campbell’s theory, abnormal experiences are only ever a consequence of the delusion; never its cause.

However, it is important to note that when Campbell talks about experience, the conception of experience to which he is referring is not equivalent to experience as it is standardly conceived by the empiricist, to the end that even though Campbell’s rationalist approach allows that in some delusions anomalous experience may perhaps be a downstream feature, the set of delusions to which this applies is actually smaller than one might think.

To explain. The theory of experience to which Campbell subscribes and so which supports his account of delusion is a relationist (disjunctivist) hypothesis (Campbell 2002). Hence, Campbell’s account of delusion can be labelled a relationist rationalist approach. The key premise of the relationist view is that perceptual experience is direct, i.e. there is no transitional interface between an objective feature and a subjective first-person experience of that feature, and so in perceptual experience one literally engages with the objects and properties of one’s surrounding world (e.g. Snowdon 1980-1; Martin 2002; Campbell 2002; Haddock and Macpherson 2008 [2011]). Yet it follows that in perception and in hallucination experience cannot be of the same kind, that is to say there is a fundamental disparity (or disjunction) between the two states (e.g. Macpherson 2013).

In brief, the relationist proposal is that in perception experience is direct and occurs in virtue of a relationship which temporarily holds between a minded perceiving subject and a perceivable object or property in his or her world. So if we take the example of Smith seeing (veridically) a tree. On the one hand, there is a tree. On the other hand, there is Smith, who is equipped with a perceptual brain. And when Smith looks at the tree, his perceptual brain attends to the object itself, i.e. to the actual tree which is out there in the world. And in accordance with the perceptual brain attending in this manner, so there is actually something-it-is-like for Smith to see that tree in the most direct and literal sense. That is to say, Smith’s experience is not something which happens indirectly in his brain via a representation in the mental stuff of consciousness,
because his experience is literally out there in the world in the actual stuff of the tree. Take away the brain’s capacity to attend, and experience disappears. Take away the tree, and experience disappears. The relationship is as direct as this.

So what about hallucination? On the face of it, the answer is simple, there is no such thing. That is, if hallucination is something which happens in the absence of an appropriate stimulus, and if an appropriate stimulus is a necessary criterion for actual perceptual experience, then whatever happens in hallucination it is categorically not an actual perceptual experience, though it could be something else (Macpherson 2013). For instance, one way in which hallucination might be accounted for is to identify the would-be experience as a consequence of the brain, and specifically the brain’s cognitive capabilities, effecting a kind of trick (ibid.). Basically, the brain’s cognitive machineries conjure up the idea of an experience and in so doing lead the subject to believe that he or she is perceiving x. Yet regardless of how deeply the subject believes they are seeing, hearing etc. the fact is that he or she is experiencing nothing at all. In other words, according to the relationist in hallucination there is never something-it-is-like, but only ever something-it-seems-like for one to *see, *hear etc anything at all (ibid). And this principle holds whether the hallucination in question is the sound of a buzz, a sensation of pain, a dream-like episode, or the seemingly perception-like sights sounds and feelings of a truculent second head which distresses one so much that in despair one considers chopping it off with axe before attempting to shoot it dead.

This sums up the primary claims made by three prevailing theories of delusion and their respective underlying assumptions re experience. However, there are two other important categories of assumptions which in one way or another play a key role in each the prevailing views. Both have been mentioned previously, but for the sake of cohesion I recap them here. The first category of assumption concerns rationality, or its converse irrationality. And the second concerns the idea of an underlying ‘organic’ cause.

Rationality and pathology of the brain
The assumptions made by the prevailing theories with regard to rationality fall into one of two groups. The first takes any subject who suffers from delusion to be a fundamentally rational agent. That is, all things being equal, there is held to be nothing deeply amiss with his or her capacity for reasoning. This is the viewpoint assumed by the empiricist one-factor thesis (e.g. Maher 1974). The second way of thinking about rationality is to say that in any case of delusion there is, by necessity, something deeply
amiss re the subject’s (or more accurately, his or her system’s or brain’s) capacity for rational processing. That is to say, the subject is held to be a deeply irrational agent. This is the viewpoint assumed by the empiricist two-factor theorist (e.g. Coltheart et al. 2011) and Campbell’s relationist rationalist account (Campbell 2001).

Finally, the assumption of the fact of an underlying ‘organic’ cause is a point about which all of the prevailing theories agree to the extent that each theory maintains that the anomalous factor(s) it identifies as having explanatory significance is (are) in every case the effect of an underlying pathological and specifically neuropathological cause. Thus for the empiricist one-factor theorist, in any given case of delusion, there is always a discrete neuropathology which causes the explanatorily significant anomalous experiences a subject with delusion necessarily has (e.g. Maher 1974 pp. 100-2). In contrast, Campbell’s relationist rationalist thesis claims that by necessity in any given case of delusion a discrete neuropathology causes the anomaly present in the foundational propositions of the subject’s brain (Campbell 2001 pp. 97-8). And regarding the empiricist two-factor account, the thesis is that in any case of delusion, there must be two discrete neuropathologies which are jointly necessary but individually insufficient to give rise to a case of delusion (Coltheart et al. 2011 pp. 282-88). One of the neuropathologies impacts on the way in which experience is processed, and so underlies the first explanatory factor of anomalous experience. The second neuropathology affects the cognitive system, and so causes the second factor of abnormal cognitive processing (ibid.).

I say more about all of the above assumptions when I discuss the prevailing approaches to delusion in chapter two. However, even at this early stage of explanation it might be thought that the different approaches seem to present with different strengths and weaknesses. For instance, the empiricist approaches with a necessary appeal made to actual experience might look well positioned to explain delusions which have an experiential element, whereas Campbell’s relationist rationalist approach which by necessity eschews abnormal experience, and instead accounts for delusion in terms of abnormal cognition alone, could seem well positioned to explain delusions with zero experiential factor. However, the converse of this is that there is no clear reason to think that any of the prevailing theories can meet every kind of case – and indeed closer examination reveals that matters are not even as clear-cut as they might at first seem. In other words, when faced with the task of accounting for delusion as delusion is
recognised to occur in a clinical setting each of the prevailing theories of delusion runs into specific difficulties as regards some or all of the assumptions outlined above, which indicates that as things stand there could possibly be something of a stalemate in place. Therefore in light of this prospect it seems not unreasonable to suggest approaching the problem of delusion somewhat differently, and the way I propose doing this is from the perspective of a Dennettian plus Prediction Error account.

A Dennettian plus Prediction Error Approach

The approach which I propose to develop and apply to the problem of delusion is a joining together of two discrete theories. The first is Daniel Dennett’s theory of mind, qua intentional content and consciousness (e.g. Dennett 1987 [1989a], 1991 [1993]). The second is Prediction Error theory, which defines a set of rules concerning how within a system, qua a living brain, information is selected for use (e.g. Friston et al. 2006). I refer to the combined view as a Dennettian plus Prediction Error (D-PE) account. However, given that Dennettian theory and Prediction Error theory are discrete and dissimilar theses, it might be questioned, what are my reasons for joining them here? In response I would say that while each theory is independent and like all theories comes with its own strengths and weaknesses, combining the approaches into a unified D-PE account provides a way for each side of the model to tell a more comprehensive story then it is able to tell on its own.

For instance, on the one hand, the fine grained design-level focus of prediction error reporting offers a way of describing what is going on inside the information sorting pathways of a wide Dennettian system in both typical and atypical cases, which is interesting because atypicality is not something which Dennett’s highly normative intentional model is ideally equipped to address in detail. On the other hand, Dennett’s ideas regarding a wide intentional system give Prediction Error theory a means of addressing aspects of the human condition, such as desire, or love, or just having a generally good or lousy day, which of itself prediction error reporting inside a brain does not seem the best tool to use to discuss.

Thus in this section I introduce both sides of a Dennettian and prediction error view. First, I outline Dennett’s theory of mind, i.e. the core principles of what Dennett says re his own thesis. Second, I outline the principles of prediction error reporting as prediction error theorists apply the theory to the functional mechanisms of a living
brain. Third I highlight some of the consequences of pushing together the two theses. Then, later, in chapter three, I combine the two theories into a unified D-PE account and in so doing develop a model which I apply to the abnormal test cases of delusion in chapter four.

Daniel Dennett’s Theory of Mind

Daniel Dennett’s theory of mind establishes the essential framework in which a D-PE model survives. Typically it is approached in one of two ways. One way is via Dennett’s account of content, i.e. *The Intentional Stance* (Dennett 1987 [1989a]). The other way is via Dennett’s account of experience, i.e. *The Multiple Drafts* model of consciousness (Dennett 1991 [1993]), which in later works Dennett refers to as *Fame in the Brain* (Dennett 2005 [2006b]). It is incorrect, however, to think of Dennett’s approaches to content and consciousness as fundamentally independent. They are not. Instead the Dennettian view is that of a well-designed adaptive system which is wide, complex, and functions in parallel at different levels, and within which system content (information) is the primary criterion which in some cases results in a conscious (subjective, first-person) experience of the world (e.g. Dennett 1991 [1993] pp. 457-8).

Thus, given that the purpose of this thesis is not merely to explicate the two familiar sides of Dennettian theory, but to take the of idea of a Dennettian system, insert the dynamics of prediction error reporting into it, and then examine what (if anything) of value the combined model can tell us about abnormal cases of delusion – which by definition involve dodgy content, and some of which involve markedly abnormal experiences – so I propose to approach the explanation of Dennett’s theory from the bottom-up. First, by outlining the foundational principles of a Dennettian system, i.e. what it is and how the model along with its relevant terminologies fits together. Then by considering how Dennett’s familiar accounts of content and consciousness fit in. It is an approach which is not without president, as it reflects the explanatory line which Dennett takes in much of his own work (e.g. Dennett 1991 [1993], 1996, 1998 [1998a]).

*The Dennettian system*

Dennett’s theory of mind rests on the idea of a single, wide, multi-layered or ‘massively parallel’ system, which for explicatory purposes can be divvied up three ways. First, the
beating heart of the system is \textit{a living human being}, this is the person or subject. Second, inside the human being, at the operational hub of its bodily machineries, is \textit{a living human brain}. And third, outside the person, is \textit{a surrounding environment}, which the person (or alternatively, the living bits of the system) must negotiate successfully-enough if he or she (or they, or it) is to survive (e.g. Dennett 1991 [1993] pp. 171-82).

\textbf{Good design and basic requirements}

According to Dennett, survival and, by extension, replication (which is roughly synonymous with reproduction, but only partly synonymous with sexual reproduction) are the foundational requirements of any living system and as such the need to address these requisites motivates everything the living human system subsequently does. To explain how this works, Dennett makes significant appeal to the idea that in any circumstance the entire system functions as it ought to according to the rules of good design (e.g. Dennett 1987 [1989b] 17, 19-20, 1987 [1989c] pp. 49-50), by which Dennett means evolutionary design via the pressures of natural selection (Dennett 1991 [1993], 2017), which in the context at point Dennett maintains occupies two interlocking roles.

The first role – or rules of design the living system is bound by – are those which it has acquired via its long history of biological evolution (e.g. Dennett 1991 [1993] pp. 171-226). These features are fixed or ‘hardwired’ into the system’s biology and biological workings, and they present as a closed set of typical responses to a limited set of typical criteria (pp. 173-82). The system’s hardwired response to grab at likely boons or withdraw from likely harms as these appear in its local environment is an example of this feature of its design (pp. 176-82).

The second role – or rules of design a living human system is bound by – are those which it has acquired via the continuing pressures of social evolution (pp. 182-252). These are the acquired, learned or ‘softwired’ dispositions to behave in certain ways re certain circumstances which individuals within a population replicate (mimic, share, copy) between themselves (pp. 182-7, 199-208, 209-226). Cultural, social and familial mores concerning criteria such as relationships, language, entertainment and the latest must-have items all fall into this wide and flexible group of motivating stimuli along with the predispositions to react to them in characteristic ways (pp. 199-208).

Nevertheless, behind the system’s newer-fangled variable drives for novel criteria, the old foundational requirements of survival and replication are still at work, with the difference that as the system evolves increasingly sophisticated (top-end,
higher-order, cognitive) capacities, its basic (bottom-end, brute) requirements start to be addressed not only via hardwired biological rules, but also by means of stuff in its wider, e.g. social, environment deemed (it seems not always prudently) likely to satisfy the same old foundational needs (Dennett 1987 [1989c] pp. 48-50, 1991 [1993] p. 181).

In other words, the same old bottom-up drives continue supplying the energy to meet the system’s newer ‘fancier’ objectives. However, corresponding to the shifting downstream of the system’s objectives, so motivational energy tends to become similarly removed (e.g. Dennett 1991 [1993] 173-82, 187-193). Thus brute grab and withdraw responses attenuate to needing, desiring, disliking, disapproving etc. So for an adult human being upstream imperatives towards survival by means of, say, food, mobility and basic interaction, might be rewritten as a desire to lunch at a pleasant restaurant with an old friend (cf. Dennett 1987 [1989b] p. 20).

**A rational and virtuous system**

So how do all of these bottom-end and higher-end behaviours come to pass? The answer is in a deeply rational manner. That is to say, at any moment and in any set of circumstances, the entire system is striving to behave precisely how it ought to behave (by design) so as to address its requirements (needs, wants, desires etc.) in a satisfactory way (Dennett 1987 [1989b, 1989c]). As Dennett observes:

[N]ature has built us to do things right; look for systems to believe the truth and love the good ... (Dennett 1987 [1989c] p. 50.)

This is an optimistic hypothesis. And Dennett explains it by the idea that at every stage of activity across the massively complex and connected living system specialist organs and utilities are primed (by design) to recognise and extract information concerning relevant criteria, i.e. patterns or norms, from a background of informational noise (e.g. Dennett 1987 [1989b], 1991, 1991 [1993] pp. 221, 265 etc.). Thus, all things being equal, the patterns selected-in are going to be those which (by design) are meaningful, i.e. important and valuable, to the system, while those which are not deemed meaningful will be selected-out or left behind.

This is not to suggest that every pattern which succeeds at being selected will be a delightful pattern from the perspective of an individual subject – and especially not so from the perspective of an individual personality – for the reason that the dynamics at
point are blind and play out via a highly competitive process across time and on the whole throughout systems, such as organs, persons and populations (Dawkins 1976 [2016]; Dennett 1991 [1993] pp. 193-208). Thus, one, accidents may happen, and errors can befall. And, two, (as most adults have discovered) some things which are ‘right’ aren’t particularly sweet, whereas (as all children need to discover) many things which appear sweet aren’t particularly ‘right’ (cf. Dennett 1987 [1989c]). Accordingly, within this model it can be predicted that mistakes are going to happen. However, for the most part, a particular system doesn’t need to be infallible, it only needs to get its selections right-enough for enough of the time (Dennett 1987 [1989b], 1999).

(Basic) Belief

Concerning the selection of patterns. The way this is held to work is that in the living body there are specialist organs which are primed by design to look out into the surrounding environment until a likely bit of pattern chances by, whereupon the pertinent bit of pattern will (perhaps) be selected-in by the organ (e.g. Dennett 1991 [1993] pp. 55-60, 112-3, 176-82, etc.). That is to say, the information contained in the (bit of) pattern gets dragged inwards, into the information-sorting pathways of the living system, and (potentially, in one form or another) all the way in to problem-solving organ at the hub of the living system, the living brain.

However, a result of its being selected is that the integrity of the pattern fractures. That is, the moment a (bit of) pattern is selected-in to the living body, the information composing the pattern breaks up across (biological, neural, cerebral) space and time as it is incorporated into the massively connected network of data which at any given time is streaming at different speeds and via different routes throughout the information-sorting pathways of the body and brain (e.g. Dennett and Kinsbourne 1992 pp. 185-234) in an ever-changing environment of shifting data.

And dispersed across this environment, the informational contents of the informational pathways are incessantly realising and re-realising something like an ever-changing kaleidoscope of basic informational patterns (e.g. Dennett 1991 pp. 44-5), where any particular pattern must compete against other instances of shifting patterns if it is to survive (e.g. Dennett 1991/1998 [1998b]).

And within this shifting melee of basic informational content, any particular instance of an informational pattern is what Dennett calls belief, or sometimes basic belief in order to differentiate the idea of belief as a shifting pattern of rudimentary data from
the more conventional conception of belief as a sophisticated linguistic or propositional state (ibid).

Dennett explains and argues for his conception of belief in many places (e.g. 1978 [1981b], 1987 [1989a], 1991, 1991/1998[1998b]). Yet, in summary, four key points are that:

1. **Beliefs are non-conscious.** Belief in the Dennettian sense refers to shifting patterns of rudimentary informational content as this variously moves around, combines, transacts, transfers etc. across the informational pathways of the entire system, and specifically those of the body and brain. It is content of a sort that a personality could not ever be aware of (e.g. Dennett 1978 [1981b], 1991). However, belief content is not only located inside a body. This is because much of the informational pattern is ‘out there’ in the environment itself. Moreover, there is a division of belief which is not actually located anywhere and with regard to which Dennett makes a comparison to the contents of a virtual library (Dennett 1991 [1993] pp. 359-62). That is to say, there is a vast field of potential from which information can (perhaps) be extracted *if the need to do so arises*, for example if one is prompted to recall a far distant memory or imagine what it is like to take a trip to the moon (Dennett 1987 [1989c], pp. 55-6, 70; 1991 [1993] pp. 441-54).

2. **Beliefs are noisy.** Patterns of belief involve countless moving parts working at different speeds, along different pathways, in different specialist areas, with different local objectives, so that at any one time the body of belief presents as a big shifting jumble requiring the processing capacities of a highly motivated living brain to sort out the story, which it does by imposing (in significant part, linguistic) concepts at pertinent places onto the ‘multidimensional’ fundamentally non-linguistic contents of basic belief (Dennett 1991 pp. 44-6, 1998b). As Dennett says, in this regard, the brain collects true ‘sentences’ (Dennett 1978 [1981b] pp. 304, 306, cf. Dennett 1995 [1998c] pp. 328-9). I say more about this re ‘judgements’ in a moment.

3. **Beliefs are unstable.** The context in which beliefs exist is one of chronic instability, since old content is forever being tested and amended against new and revised information at different speeds and in different (specialist) areas thus effecting an ever-changing ‘snarl’ of basic information in a system which is obsessively driven by a need to work out what

4. **Beliefs are impotent, for practical purposes.** The patterns of information contained in beliefs have a potential to inform the wider system about what is (likely) happening in the world and thereby instruct it what to do next by way of triggering pertinent patterns of downstream behaviour. However, so long as a pattern remains in circulation in the role of belief it is moving far too fast and is far too unstable to be of much practical use. As Dennett says, beliefs need to be thought of as ‘complex behavior-disposing organs’ (Dennett 1991 pp. 45-6), but other criteria have to be in place before the energy and information contained in a belief can have any actual effect (e.g. Dennett 1978 [1981b], 1991, 1991 [1993] pp. 111-3, 135-6 etc.).

**Judgement**

The context so far is that of a deep rich ever-changing melee of basic belief. However, from out of this turmoil of information something must be selected, simply because *something must be selected*, this is how a living system works. In other words, repeatedly and across itself the motivated brain needs to opt for this pattern of belief over that pattern of belief as a way of moving the living system forwards, and it can’t select at random but it must do so pointedly in a way which (likely) maintains the integrity of the system. That is, the heart needs to beat, and the lungs need to work, at precisely the same time as all of the other behavioural, experiential and cognitive stuff is going on.

How does the brain manage this task? Dennett proposes that it happens via a process involving repeated rounds of processing in which competing patterns of shifting content are tested and retested following a logic which is not rigidly mathematical but is ‘at best heuristic’ until at some point the brain stops its deliberations and selects, chooses, opts for a particular pattern of belief, or as Dennett says, the brain ‘bets on the truth’ (e.g. Dennett 1981b pp. 303, 304, 306). The idea is that the brain places its bets by effectively applying or imposing a pertinent top-down concept onto the shifting bottom-up belief set which works to corral or fix a relevant pattern of content and so effectively slows it or stops it for a time in its tracks (e.g. Dennett 1978 [1981b], 1991, 1991 [1993] e.g. pp. 333-5, 1992).

point about judgements is that, at any time in the brain there is not one single judgement, nor even one single type of judgement in play. Instead, there are very many different judgements in very many different locations (e.g. Dennett 1991 [1993] p. 113), extending from those which have meaning in the brute roots of biology, e.g. judgements about the presence of \(O_2\) as opposed to \(CO_2\) molecules and judgements about typical patterns light and dark, to those which have meaning at the levels of higher-order functioning and involve concepts which are highly linguistic in nature and which work to conceive a model of culture a person lives within e.g. ‘the concept of a nickel-plated combination corkscrew and staple remover’ (Dennett 1991 [1993] pp. 187-93).

Moreover, different judgements are motivated by different degrees of urgency, with some being realised very fast and others far, far slower (ibid.).

And always various judgements will be at various stages of processing or ‘drafting’ (e.g. Dennett 1991 [1993] p. 113).

Yet stick the miss-mash of judgments together and it realises the ever-changing story which a particular brain is continually telling itself about the (likely) circumstances of its changing environment. In Dennett’s terminology, this is something like a narrative stream (ibid.).

Four other relevant points about judgements in the Dennettian sense are that:

1. **Judgements are not beliefs.** Although judgements have their roots in belief, they are not beliefs. For instance, contrary to the wide, changing, multidimensional nature of belief, judgements have definite edges which are potentially expressible, plus (in theory) in a brain judgements are ‘precisely locatable’ in space and time (Dennett 1978 [1981b], 1991, 1991 [1993] e.g. p. 113, 1998).

2. **There is no necessary (one-to-one) correlation between an item appearing in the environment and content re that item featuring in the narrative stream.** The nature of the process is one of progressive change and selection, where at any stage what gets selected-in is in part motivated by need and in part marked out by expectation, and so a stimulus being in the environment is no guarantee that information connected to that stimulus will be selected-in, and continue to be selected-in, in whole, in part, or in any particular form (e.g. Dennett 1991 [1993]; Dennett and Kinsbourne 1992).
3. *Judgements are only ever relatively fixed.* The content of the story which the brain is continually telling itself is only ‘rather like’ narrative, in part because of the motley nature of the different types, speeds etc. of its contents, and also because these contents are always on the move. That is to say, the narrative stream and the judgements which comprise it are never static, but always in a process of ongoing revision. This is the principle supporting Dennett’s idea of *multiple drafts* (e.g. Dennett 1991 [1993] pp. 17, 111-43, 170 etc.)

4. *Judgements are not experienced.* The narrative stream is the brain’s story, not the personality’s story. That is to say, you and I are oblivious to the shifting patterns of content-fixation which are continually (and figuratively speaking) flowing around the surface of our respective brains (Dennett 1991 [1993] pp. 113, 333-8, 365). For one thing, many parts of the narrative are simply nothing to do with consciousness. And for another thing, the role of the narrative stream is not experience, but moment-by-moment it works to prime or ‘ready’ the brain with regard to the (potential, prospective) realising or releasing of pertinent patterns of behaviour – including those behaviours leading to experience – should circumstances demand it. That is to say, if and when the brain is prompted to release the contents of its narrative in meaningful places, which happens in consequence to what Dennett calls *probes.*

**Probes**

A probe (prompt, trigger) is essentially a demand for action which is imposed on a part of the brain’s narrative, and so prompts the release into behaviour(s) of the content (i.e. pattern of energy and information) held at that specific point. Figuratively speaking, a prompt pulls open the gate of the conceptual corral. And prompts can be of different kinds. There are environmental triggers; for instance, the onset of rain might prompt one to run for shelter or run outside. There are associations; for instance, complete the phrase: *a cup of ___*. There are imperatives; for instance, ‘LOOK!’ There are nudges; for instance, ‘You might want to think about x.’ And there are questions, which may be a question that another subject asks of us or a question which we ask ourselves, and which partly through association and partly through the mores of convention demands an apposite response, even when such a response is not straightforward. For instance:

‘I’d like some baked beans, please.’
‘Yes sir. How many?’
(Dennett 1987 [1989b] p. 20-1.)

However, and importantly, it is a fallacy to think that there is ever one single or even one optimal response to any given prompt. This is because, given the way in which the brain’s narrative stream is always on the move, so a different prompt at the same time or the same prompt at a different time or in different circumstances might yield a markedly different but still equally valid response (e.g. Dennett 1991 pp. 45-8, 1991 [1993] pp. 113, 135-8, 143 etc).

Behaviour

At any and every pass there must be a collapse into one behaviour or another behaviour, again for the reason that this is just how a living system works. The collapse might be decisive, or it might not, and it might trigger an observable pattern of behaviour, or it might not, all of this depends on other factors. Nevertheless, a motivated system always needs to do something, or more accurately do very, very many somethings, and at any moment these somethings might be (any combination of) behaviours which are fast, slow, functional, motor, linguistic (thought or speech) or experiential in kind.

Two significant points about behaviour are that:

1. There is no direct correlation between the onset of an item of content-fixation in the brain’s narrative stream and the realisation of that content in behaviour, including experience. As Dennett says re experience, ‘It is always an open question whether any particular content thus discriminated will eventually appear as an element in conscious experience’ (Dennett 1991 [1993] p. 113). In part, this is because at a particular time circumstances may not warrant the collapsing of a particular portion of the brain’s narrative into behaviour. And in part, it is because of the way in which the contents of the narrative stream are under continual revision, so that in whole or in part a pattern of content-fixation may be amended many times over before it is ever in a position of being fulfilled in behaviour – indeed, if it ever is in the position of being fulfilled in behaviour at all. Dennett approaches this feature of his theory most explicitly as regards experience, and the way in which contents are always subject to the influences of redrafting (Dennett 1991 [1993] pp. 115-26; Dennett and Kinsbourne 1992), and it establishes the nub of the multiple drafts model of consciousness (ibid.).
2. Just because behaviour (of any stripe, including experience) is performed does not guarantee that it is going to be noticed. At any moment in the context of any living system and/or any collection or population of living systems countless behaviours are being performed. Many of these behaviours will be invisible to everyday observation. Many will be potentially observable, but either never observed, or if observed immediately ignored or forgotten, say, because they are trivial, or unexpected, or unintelligible, or because the observer is hampered by anxiety, confusion, preoccupation and so on. Nevertheless, in all of these cases, the behaviours still happen and their patterns are still real patterns in the world, only they are real patterns which never succeed in selection from the medium of (environmental) noise in which they exist (e.g. Dennett 1987 [1989b], 1991). However, clearly, some patterns of behaviour do get selected and interpreted and acted on. And we can say that this happens either centrally (personally, internally) i.e. when a living system picks out patterns in its own behaviours, or non-centrally (adjacently, distantly, externally), i.e. when a living system picks out patterns performed by another system in its locale. And of the patterns selected in this way, the contents of some might succeed for long enough and relatively unchanged enough and in circumstances conducive enough for downstream behaviours including experience in a genuine sense to be realised, for instance, for Betsy to become aware of the thimble she is looking at (Dennett 1991 [1993] pp. 333-8), and this is the nub of Dennett's idea of consciousness qua fame in the brain (Dennett 2005 [2006b]).

As for the question, why is it that some patterns and not others are extracted and observed? The answer is because, every observer (subject, his or her system) is bound by the same old motivational pressures to understand (adjust to, learn about, predict) the goings on in the world around them and in so doing act in a way which maintains its (his or her) basic requirements. Thus the patterns selected are those which across time and on the whole ought to (i.e. have tended to) provide for the satisfaction of the system’s needs (wants, desires) – which allows for simple interest, which is to say, a highly motivated ‘epistemically hungry’ system can make for an ‘epistemically curious’ subject, so that some patterns may be attended to simply because they are novel, or pretty, or whatever. And always there is the caveat that within this framework accidents and glitches and miscalculations can and likely will happen, so there is always room for chances or coincidences or curiosities to occur.
This sums up the outline of the Dennettian framework re a self-motivated living system as it engages in a continual process of interaction and adaptation within itself and its environment via the repeated and massively parallel selection and testing of pertinent patterns of information from a background of noise with the overall intention of maintaining its own integrity within an uncertain world. And it is from in this framework that Dennett reflects on two conventionally significant ways in which human observers, and specifically modern adult observers from within our own culture, interpret certain behaviours which are recognised as happening at a level of mentality, i.e. as mental states. The first concerns the behaviours and attributions of intentional content (i.e. states such as belief, desire, wishing, fearing etc.), and it is this part of the theory which Dennett presents as The Intentional Stance (Dennett 1987 [1989]). The second concerns the behaviours and attributions of consciousness (awareness, experience), and it is this which Dennett presents as The Multiple Drafts Model of Consciousness (Dennett 1991 [1993]).

Dennett’s Theory of Intentional Content

Dennett’s theory of intentional content is the idea that driven by the need to understand what is going on and therein what to do next, a subject (or more correctly, a system) is constantly interacting with its surrounding environment, and in doing so, encountering others (systems, subjects, objects) at work and the patterns of behaviour with which they present. Thus in an effort to assimilate, integrate and if necessary adapt to the shifting situation, the subject (system) depends on its nature, biography, knowledge, desires etc. to organise the flow of data and in so doing extract meaningful patterns of information from the melee of environmental noise, and in some cases these patterns might be meaningful in a sense that we, modern humans, have learned to conceptualise using intentional terms, such as belief, desire, fear, anxiety, joy, knowledge, imagination, awareness and so on (Dennett 1987 [1989a] p. 7, 1991/1998 [1998b]). Thus in certain circumstances Smith might look into himself and conclude that he feels hungry and wants lunch, then look over at Jones and infer that he (Jones) must be feeling the same way, thinking the same thoughts, and likely to be hankering for some beans on toast about now.

Dennett proposes that to engage with the world at this level of interpretation is to adopt an intentional strategy or intentional stance towards the object or objects at point
(Dennett 1987 [1989b]). Why do we do it? The answer is because within our own population across time and circumstances the concepts (patterns, norms) at point have proved themselves to be a good-enough shorthand means of predicting and interacting with other systems, and specifically, with other minded agents in a way which affords a heads-up about what is (likely) happening now and therein what should (likely) happen next (ibid). In other words, via adopting an intentional strategy Smith can feel safe proposing the notion of lunchtime to Jones.

However, it needs to be stressed that within the Dennettian model what is being ascribed (i.e. the feeling, belief, desire, etc.) is ascribed in virtue of a pattern of information which is set in a shifting composite of conditions involving patterns of behaviours performed by a living agent in a certain set of circumstances and which composite necessarily includes the presence of the observer of the pattern himself or herself. Accordingly a belief, desire etc. cannot be construed as identical to an actual state of brain, or even to an actual state of body and brain. Instead, what we term belief (desire etc.) is a complex multilevel pattern, which from an ontological perspective is real enough in that the pattern is actually there in the environment waiting to be selected when the circumstances are right (Dennett 1987 [1989b], 1991). Nevertheless, as Dennett says, this is a ‘mild’ type of realism (Dennett 1991), not only because in any single case is it effectively impossible to demarcate the actual pattern of the belief (desire etc.) – as opposed to the observer’s judgement re the pattern – but also because in no two cases involving the same ascription can the patterns underlying it be precisely the same. As Dennett says, the ascription of belief (desire etc.) happens when a wide deep multidimensional pattern of information gets filtered through the one-dimensional concepts of the speaker’s (ascriber’s, observer’s) natural language (p. 45).

*Dennett’s Approach to Consciousness*

Dennett’s theory of consciousness also rests on the notion of a living system (subject) that is located in an environment within which it is motivated to act and interact in order to satisfy its requirements, and which (system) negotiates this task via the deployment of specialist organs and functions primed (by design) with expectation, and which (specialists) are continually skimming and scanning their surroundings for (bits of) pattern relating to various meaningful criteria, and in accordance with which the system works to derive a meaningful story about what is going on in its world, and in
relation to parts of which it (the system) might be motivated to act (Dennett 1991 [1993], Part II An Empirical Theory of Mind, pp. 101-282).

Again as observed previously, many of the brain’s judgements which comprise the story will be fast spontaneous and relate to brute function, while others could be far slower and involve highly intellectualised themes. However, at a stage somewhere between bottom-end function and top-end cognition, Dennett argues that there is a special class of judgements to which he gives the name presentiments (Dennett 1991 [1993] pp. 113, 343, 364-5, 457). And presentiments are instances of content-fixation within the brain’s narrative which presume to inform the brain about pertinent objects and properties in the system’s qua subject’s ‘observable’ world.

A presentiment can be described as what happens when a likely bit of bottom-up pattern of a certain sort is corralled in the brain's narrative by a top-down expectation of a certain sort. Like all judgements, presentiments have conceptual edges, are derived in multiple different places across the brain, are propelled by different strengths of urgency, are at various stages of drafting, and are inaccessible to personal experience, since their function is to inform the brain about what is going on ‘out there’ (Dennett 1991 [1993] pp. 113, 365). For instance, they could be informing the brain that *Psst, look out, there's a big hole in the ground up ahead.*

However, among judgements what makes presentiments special is that the concepts which corral them are of a kind which, if prompted, trigger downstream behaviours of a kind about which the subject, at the level of the personality, could (perhaps) be aware, i.e. experience, reflect on, argue about their content etc. And specifically in this role, Dennett singles out the concepts supplied by our words, namely those of a speaker’s natural language (Dennett 1991 [1993] pp. 225, 300-1, 446-7 etc). The idea being that when bottom-up data is corralled by the kinds of concepts which words afford then definite (conceptual) edges are imposed on the shifting and uncertain noise, which leads to a sort of (fragmented, messy, shifting, phenomenological) text being generated across the brain’s story, which, if prompted in the right way and at the right time, releases downstream behaviours which (potentially) move selected information in the pattern to the level of the personality, whereupon via an apposite mode of fluid informational expression the subject may learn about, that is *see*, a big hole in the ground (e.g. Dennett 1991 [1993], 1993b, 2002a, 2003, 2005 [2006b], 2007).

It might be queried in a design based model what is the supposed advantage of having information present in a mode that the subject (as opposed to his or her brain
alone) can engage with? And following Dennett, I suggest that the answer is because it increases the system’s options to seek out the good stuff and avoid the bad (e.g. Dennett 1984 [2002b], 2003 [2004]). Essentially, pulling content upwards out of the mechanical workings of brute biology affords the system room for manoeuvre, i.e. more time, space and capacity for choosing, so as to satisfy its requirements in novel and interesting ways. And being able to see, hear, feel, think about etc. the stuff of the world can (potentially) be of advantage with this.

It might also be queried, is that it!? What about the phenomenology of consciousness? And the answer is that, there is no such thing (e.g. Dennett 1991 [1993] pp. 365-7, 1993b; Dennett and Kinsbourne 1992). That is to say, when you or I or Smith or Jones look out and witness the greens and blues of the fields and sky at no point in proceedings do any of us consume (have, undergo, enjoy) an experience of actual consciousness resplendent with phenomenal contents or character. Instead, there is just a lot of shifting data and ever-changing judgements about greens and blues and fields and sky across a channel of exchange via which it truly seems to you and me and Smith and Jones that we are seeing (hearing, touching etc.) the patchwork of shapes and colours which are out there in the world.

This almost sums up Dennett’s theory of consciousness. Although I say more about it when I combine Dennettian theory with the principles of Prediction Error theory and then apply it to the cases of delusion in chapters three and four. However, before finishing this outline it may be observed that although Dennett’s approach to consciousness is very different to the relationist theory of perceptual experience, there are some superficial similarities between the two accounts. First, consistent with Dennett’s model of consciousness and the relationist’s explanation of the bad case of hallucination, there is common ground in that both theses maintain that there is never something-it-is-like for one to have said experience, but only ever something-it-seems-like to see, hear, feel whatever is at point. Second, and again not entirely dissimilar to the relationist’s bad case, according to Dennett’s model, what it seems like to experience anything in any case bears no necessary (one-to-one, direct) relation to actual objects and properties in the world. This is because across the informational pathways of the system beliefs, judgements and behaviours are held to be broken, choppy and always on the move, and likewise the nature of experience is that of a broken, choppy, ever-changing story of content which in everyday circumstances works sufficiently well to
convey an illusion of objective unity, but which illusion is simply a sign that the system and problem-solving brain is doing its job.

Prediction Error Theory

Prediction error theory as it applies to the workings of the living brain captures the idea that one important way in which a brain processes information re perception and learning is by using the epistemic tension between the patterns (of data) it expects to find circulating its environment (i.e. its predictions) and the noise in the data with which it is presented at any one time (i.e. predictive error) to motivate adapting its behaviours so as to minimise error (and so maximise prediction), and that it does this by means of an ongoing process of testing the disparity between expectation and noise (i.e. prediction error reporting) in concert across many parallel stages of its functional machineries, and in so doing inferring conclusion after conclusion about the ever-changing conditions of its (internal and external) world (e.g. Friston et al. 2006; Hohwy 2013).

One of the foremost advocates of the approach is neuroscientist Karl Friston (e.g. Friston et al. 2006; Friston 2009). Yet according to Friston, the idea of the brain qua a self regulating ‘inference machine’ is not an uncommon view (Friston et al. 2006 p. 70).

For instance, Friston talks about the idea of a free energy principle for the brain (essentially, free energy being prediction error) which the system works to minimise by means of adapting itself in ways which affect changes in its actual relationships to the stuff of its environment or changes in the patterns it picks out as meaningful from its environment, and in so doing realises pertinent changes of action, perception and/or thought (Friston et al. 2006; Friston 2009). Relatedly, Bromberg-Martin et al. appeal to prediction error in the brain in connection to the role of dopamine neurones and motivational control, and specifically in connection to the encoding of motivational value and motivational salience regarding positive, negative and alerting events and the downstream behaviours this effects (Bromberg-Martin et al. 2010).

Then again, cognitive scientist and philosopher Andy Clark examines predictive processing in relation to the situated agent (Clark 2013), and the interconnection of language, cognition and perception (Lupyan and Clark 2015), and personal choice (Clark 2020).

Alternatively Jakob Hohwy (2013) presents a philosophical theory of perception which proposes that a brain working to minimise prediction error derives inferences
across a hierarchy of functional levels some of which contribute to generating a subject’s rich experiences – that is actual phenomenal experiences which the brain presents as a unified picture rendered in something like a ‘vibrant mental paint’ – of the various objects and properties in his or her perceivable world (pp. 25-6, 101-15).

And as regards psychosis, Fletcher and Frith think through the role of predictive mechanisms in the brain to explain the positive symptoms of schizophrenia (Fletcher and Frith 2009). Corlett et al. apply the principle of prediction error minimisation to explain the occurrence symptoms paralleling those of psychosis in cases involving drug use and/or sensory deprivation (Corlett et al. 2009). And Powers et al. explore the top down effect of prediction error minimisation on cases of psychosis involving perceptual hallucination (Powers et al. 2016).

Thus, as Friston observes, the brain as an inference machine, or more specifically as an organic system which derives at least some conclusions about its environment via a process which is comparable to the idea of prediction error minimisation, is an idea which is applied quite broadly. Yet despite the broadness of its application, prediction error theory – at least as it is approached here, which essentially makes use of ideas respectively developed by Friston and Hohwy, with further acknowledgment to Bromberg-Martin et al. – depends on a relatively small set of principles.

The first principle concerns the fundamental nature of the system at point, which basically means the living brain and its systems, organs and utilities, all of which are held to be arranged as per a massively connected hierarchical structure, which is both the product and ongoing instrument of selective (i.e. evolutionary) design (e.g. Friston 2009). Accordingly, Hohwy talks about the brain processing information re perceptual regularities by way of a connected functional hierarchy of low (i.e. fast, basic, bottom-end), medium, and high (i.e. slow, cognitive, top-end) levels of operation, with a fluid exchange of information passing up and down between all of the different stages (Hohwy 2013 pp. 31-2, 67-73). However, information does not simply move around in this framework. It moves in a context of addressing the requirements of a well designed self-organising system which is inherently driven to maintain its own integrity in line with conditions which sometimes change and sometimes do not. In other words:

A key aspect of biological systems is that they act upon the environment to change their position within it, or relation to it, in a way that precludes extremes.
of temperature, pressure and other external fields. By sampling or navigating the environment selectively they restrict their exchange with it within bounds that preserve their physical integrity and allow them to last longer. (Friston et al. 2006 p. 72.)

The second principle is that, when information moves around in this motivated system, it does not do so randomly, but at every stage by following a set of rules or ‘constraints on its behaviours’ (Friston et al. 2006 p. 71), the directions for which are woven deep into the fabric of the system by the ongoing pressures of evolutionary design (pp. 70-1). The upshot of this being that the complex multi-level system realises a capacity to share information between its various parts, that is to say, to ‘talk’ to itself from bottom-end to top-end and vice versa, and so work cohesively as a single structure with access to a rich informational schema which (potentially) stretches across multiple levels of processing as it works to represent the (likely) shape of the (system’s) changing world.

So thinking about the set of rules at work in the system. Prediction error theory recognises it as following a pattern which is comparable to the principles of empirical Bayes, whereupon the system (brain) is held to be continually taking leaps of epistemic faith by applying what it already knows (priors) to draw inferences (predictions) about what it isn’t so sure about (new or uncertain information) and on the basis of the feedback it receives (prediction error) gauging inferences (further predictions, conclusions) in virtue of which its behaviours adapt (e.g. Friston et al. 2006; Hohwy 2013, Part I The Mechanism, pp. 13-97). It is an environment of chronic uncertainty. As Hohwy observes, on the one hand, there is no one-to-one relation between information pertaining to an object or property in the environment and information concerning that object as it might be represented in the brain; while on the other hand, there is a potentially inexhaustible supply of hypotheses for the brain to actually choose from (Hohwy 2013 pp. 13, 15-6). Thus a brain needs a way of sorting the wheat from the chaff, i.e. of selecting meaningful patterns from noise. And the way it is proposed that a predictive brain does this is: First, the brain tests its predictions against the surrounding noise. That is, across the system and primed with expectations any particular nub of the functional hierarchy looks down or out into the churning contents of the level directly beneath it while obsessively asking one single question: Is it important? (i.e. Does it tally with my prediction?), of every bit of data that comes by. Then second, re every instance of testing the brain receives feedback in the form of one of three basic replies, and on
the basis of these acquired evidences it adapts (or not) its behaviours, and thereby the behaviours of the wider system (e.g. Friston et al. 2006; Friston 2009; Bromberg-Martin et al. 2010; Hohwy 2013). The three basic replies and responses can be characterised as follows.

1. No, it is not important. In other words, from the point of view of the brain the sample is too noisy (i.e. there is too much free energy or prediction error), so it moves to ignore that bit of would-be pattern (effectively switching off neurones to that possibility) which works to deselect the data, dropping it back into noise.

2. Yes maybe, it is important. In other words, from the point of view of the brain, the sample has enough pattern (prediction) to be interesting, have promise, but still too much noise (error) to tend a good fit. This prompts excitement around the likely pattern (effectively, the brain turns up energy to the neurones dealing with it) which keeps the pattern alive for further processing either at the same functional level or by making it available via noise to the next level up.

3. Yes, it is important. In other words, from the point of view of the brain, the sample concords with its expectations (i.e. prediction error is minimised), and with no requirement to ignore and no requirement for further testing, the energy around the pattern slows (effectively, the brain momentarily rests on its laurels), which if this continues for long enough renders the pattern open to effecting downstream behaviours of a pertinent sort.

The global effect of obsessively repeating this question and answer procedure at different levels, across different functional processes, and at different speeds and grades of urgency, is that the brain is continually and self-reliantly updating its own rich informational schema through incessantly adapting its and so the system’s wider behaviours – either by dropping or amending its predictions in line with the data and/or amending its relationship to its surroundings – in accordance with the drive to keep itself working as it ought to be working, by design.
Implications of Combining the Dennettian and Prediction Error Theses

I delay the joining together of the Dennettian and Prediction Error theses until chapter three where I explain in full my approach to combining them into a unified account. However, I conclude this introduction to the two discrete sides of the thesis by observing that there are implications to merging the two separate theories into a single theoretical account. I mention three of these implications below.

Potentially interesting consequences

The Dennettian and Prediction Error theses are two different theories, from different fields of specialism, working at different levels of explanation, and so with different concerns, and this is something that I say more about when I work to combine them in chapter three. However, it remains that there seems to be a substantial amount of shared ground between the two theories, so that should they be combined there might be a concern that together they tolerate a fit which has the prospect of being pretty unremarkable. For instance, both theses rest on the principle of a biological system which is the ongoing product of evolutionary selection, which is driven by the need to sustain its basic requirements across changing environments, and which it does by weighing its prior beliefs against ongoing inputs across a massively connected parallel framework and attending to the conclusions it draws which are thus made available to realise adaptive downstream behaviours of a pertinent sort. However, look a little closer at what is happening, and the effect of merging the theories is not as unremarkable as it might seem, since once we have the idea of prediction error reporting following its own set of rules and pulling the wires throughout the systems, structures and utilities of a wide Dennettian framework, some potentially interesting consequences start to emerge. For instance:

1. Dennett’s idea of multiple drafts is no longer limited to its role as a model of consciousness. In effect, the principles of multiple drafts have never been limited to their role in consciousness (e.g. 1978 [1981b], 1991 [1993] pp.457-9). However, as ‘multiple drafts’ gave a name to Dennett’s theory of consciousness, so its role with regard to consciousness tends to overshadow its role everywhere else. Yet when Dennettian
theory is merged with the principles of prediction error theory, it becomes clear that the dynamics of multiple drafts are in play at every single turn of this wide connected system. Essentially, the multiple drafts model of consciousness is itself redrafted into a multiple drafts model of mind.

2. The dictate of prediction error reporting is no longer limited to a role inside ‘the skull bound brain’ (Hohwy 2013 p. 15). This is because when prediction error theory is combined with Dennettian theory, the functional hierarchy gets pulled out of the skull, indeed out of the nervous system, to include in its parallel stages the workings of the wider body, and the subject himself or herself, i.e. his or her dispositions, proclivities, biography, feelings, thoughts, knowledges, experiences, dreams, longings, fears etc, and then even further out into the physical features and layers of mores and interactions at play in the subject’s environment – where at least some prediction error theorists seem to observe it at work anyway. For instance: ‘In everyday life, these rules [of inference] are applied to information obtained by sampling the world with our senses’ (Friston et al. 2006 p. 70) and are consistent with the ‘principles that govern the interrogation of scientific data’ (ibid.).

A different way of thinking about experience, Dennett’s versus Hohwy’s accounts

Dennett (1991 [1993]) and Hohwy (2013) each offer a theory of experience. And if the element of prediction error reporting is assimilated into Dennett’s theory, then both theories have the principle of a predictive brain at core. Thus it might be asked, what is the real difference between the two theses, because if Hohwy’s prediction error approach to experience is already on the table, then where is the value in introducing a Dennettian plus prediction error account?

My first response to this point would be to acknowledge that yes there are some similarities between the two approaches. For instance, both theories award a significant role to the controlling effects of prior content, as well as sharing the idea that there is no direct one-to-one correlation between the occurrence of a perceptual stimulus and informational contents relating to that stimulus being realised in the brain, but that the brain engages in an ongoing process of epistemic revision in which patterns of bottom-up data are fixed by the constraining effect of top-down concepts at different places, different speeds etc. across the brain.
Nevertheless, there are also differences. For instance, I have already spoken about the wideness of Dennett’s approach, and that he awards a significant role to the interpersonal, normative etc. relationships which exist between subjects, qua the various individuals who populate any particular environment, as contrasted with prediction-error theory’s characteristic skull-bound brain mode of approach. And even if it were to be protested that this is merely a difference in explanatory direction, that Dennett approaches matters from outside in whereas prediction error theory approaches the same matters but from inside out, I nevertheless submit that placing the bulk of the explanatory burden on the way that a human being or a population of human beings acts, copies, feels, thinks, prefers etc. presents a very different explanatory rationale to one which positions the bulk of the explanation in the design functions and failures of a biological system, qua a particular brain.

Furthermore, there is the fundamentally dissimilar way in which Hohwy and Dennett specify what experience is. For Hohwy, perceptual experience entails the subject enjoying ‘actual visual consciousness’ i.e. an actual phenomenal experience, of, say, a tree, which the brain presents via something like a ‘vibrant mental paint’ and where it is what you see which causes you to ‘issue a conceptual judgement’ that, say, this is a tree (Hohwy 2013 pp. 25-6). This is the conventional way of thinking about (actual) experience, with the difference that Hohwy is offering an account of actual experience in terms of the inferential mechanisms in the living brain. Even so, as Hohwy observes, in making the move from ‘austere’ inferences to the ‘richness’ of actual phenomenal experience, there seems to be no intuitive or straightforward step available: ‘It seems the Bayesian moves we just made would do fine if we were only interested in explaining conceptual thought about perception, rather than explaining perception itself’ (p. 26). So addressing this issue, Hohwy appeals to the notion of the system’s rich hierarchical structure and the idea of the first person perspective this affords (e.g. pp. 25, 27-34, 76-81, 116-39), which if it weren’t for the stipulation concerning actual phenomenology, in some ways does seem to overlap with the Dennettian view.

However, if we turn to consider Dennett’s approach which rejects the idea of an actual phenomenology of experience, and instead accounts for experience in terms of a story which the brain is telling itself and patterns of downstream behaviour which may result, so that in any case of (perceptual) experience there is never actually something-it-is-like for one to see, say, a tree, but only ever something-it-seems-like to see, hear, feel etc.
anything at all, then the Bayesian moves which Hohwy mentions would be wholly sufficient to tell the entire story of the subject’s rich first person experiences of world.

So yes there are substantive differences between the two theses. Yet as for which (if either) approach is empirically correct, the answer is that I really don’t know. However, when faced with uncertainty, if the judicious thing to do is not uncritically defer to but at least consider the simplest explanation, then it would seem reasonable to allow that a Dennettian plus Prediction Error approach at least deserves consideration alongside Hohwy’s handling of prediction error theory as a viable way of thinking about experience in accordance with the idea of a predictive brain.

Is a Dennettian plus Prediction Error approach empiricist, rationalist, or something else?

Returning to the problem of delusion as it was outlined at the start of this chapter, there is the question of whether a Dennettian plus Prediction Error (D-PE) approach would provide an empiricist account of delusion or a rationalist account of delusion, or maybe something else?

To recap. If the empiricist stance on delusion is that in any given case a clinically anomalous (actual) experience is by necessity the sole (i.e. the empiricist one-factor approach) or the initial (i.e. the empiricist two-factor approach) significant explanatory factor; and if the rationalist stance (and specifically, Campbell’s relationist rationalist stance) is that in any given case clinically anomalous (actual) experience is never a factor, and that deeply abnormal cognitive processing is always the sole explanatory cause, then given that a Dennettian account cannot appeal to actual experience but explains both content and consciousness in terms of cognition, so a D-PE account of delusion must in principle be a rationalist approach.

Nevertheless, a D-PE approach to delusion cannot be construed as rationalist in the sense that rationalism re delusion is standardly understood, and at best it is only rationalist in a minimal or weak sense. Why? First for the reason that within a D-PE framework the subject is approached as a deeply rational agent, which eschews the factor of deep irrationality on which rationalism as it is typically construed depends. And second for the reason that a D-PE approach happily allows that if in a case of delusion anomalous experiences happen, then these experiences could certainly play a key explanatory role. And although said experiences would not be actual experiences in the conventional sense, but experiences in the Dennettian sense of seeming to be the case, from the perspective of the subject this would not make one jot of difference when for
all the world it seems to him or her that he or she is feeling the breeze, hearing the
birdsong, or seeing the second head.

Thus as regards delusion, a Dennettian plus Prediction Error account at least
promises to offer a fresh approach.

The Organisation of the Thesis

So to look deeper into the topic of addressing the problem of delusion, the remaining
chapters of the thesis are organised as follows.

In chapter two, I examine in some depth the three prevailing theoretical
approaches to delusion, and via the three test cases of delusion specified previously –
namely, erotomanic delusion, Capgras delusion, and perceptual delusional bicephaly – I
analyse the issues they each face re the three key assumptions of neuropathology,
irrationality and experience on which each theory variously rests. The conclusion drawn
is that given the knot of difficulties faced across the cases of delusion by all three
prevailing accounts re the foundational assumptions they variously hold, it is reasonable
to consider a different way of approaching the problem of delusion.

In chapter three, I explain and develop a theoretical account combining Daniel
Dennett’s theory of mind and the dynamics of Prediction Error theory as these are
typically applied to the brain, and I do so within the context of a ‘normal’ case study,
which is actually an expanded retelling of an illustrative example provided by Dennett in
his essay ‘How to Change Your Mind’ (Dennett 1978 [1981b]). My rationale for taking
this approach is to develop a Dennettian plus Prediction Error (D-PE) model relating to
how in normal circumstances a subject makes up his mind to a specific conclusion and
then (perhaps) changes his mind in light of ongoing concerns.

In chapter four, I take the D-PE model developed in chapter three and set it
against the three test cases of delusion, which by definition are contexts in which each
subject makes up his or her mind to a specific conclusion which they resolutely maintain
despite its implausibility and reasoning and evidences to the contrary. In the discussion
which develops, it seems that the D-PE account effectively takes erotomanic delusion
and the case of Patsy in its stride. Likewise, Capgras delusion, in both the cases of Mme
M, which presents as a case calling for explanation in psychodynamic terms, and the
case of Mr A, which is a case which features substantial and significant traumatic injury
to the subject’s brain. Finally, re perceptual delusional bicephaly and the case of Mr P, I
argue that the D-PE approach not only addresses the features of the case study, but specifically re the very strong hallucinations with which Mr P presents, and from entirely within the framework of the D-PE model, a D-PE approach has the capacity to explain Mr P’s case well.

Therefore given that these cases of delusion were chosen for examination because together they offer a good cross section of the dataset of features of clinical delusion, then of the four accounts of delusion I have examined in this thesis, I conclude that the D-PE approach clearly has the widest explanatory remit, and so plausibly warrants further consideration alongside the conventional accounts.
CHAPTER 2
THE PREVAILING APPROACHES TO EXPLAINING DELUSION

Introduction

In chapter one, I introduced the topic of the thesis as delusion as it presents in the clinical field. Then I identified the problem of delusion to be addressed in the thesis as determining which philosophical theory of delusion best meets the data set of features associated with clinical delusions, and thereby specifying which of the recognised explanatory models of delusion works best.

The two explanatory models I consider are, on the one hand, empiricism re delusion, or the idea that in any case of delusion by necessity anomalous experience is the key explanatory cause, and on the other hand, rationalism re delusion, or the idea that in any case of delusion by necessity the key explanatory factor is anomalous cognitive functioning alone. And to identify which of the models works best as regards the cases of delusion I consider, I elected to examine three established theoretical accounts. The accounts are: the empiricist one-factor theory of delusion, which maintains that in any case of delusion abnormal experience is always the sole explanatory cause (e.g. Maher 1974); the empiricist two-factor theory of delusion, which maintains that in any case of delusion abnormal experience is always the primary explanatory factor, but it is always accompanied by abnormal cognitive processing (e.g. Coltheart et al. 2011); and John Campbell’s rationalist theory of delusion, which maintains that in any case of delusion abnormal cognitive processing is always the sole explanatory factor, and (markedly) abnormal experience never assumes a significant explanatory role (Campbell 2001).

So now, in chapter two, I begin my analysis of the problem. And I start by setting out the arguments put forward by each of three theories, with the aim of identifying more clearly what each theory relies on and claims. Following this I organise my discussion of the theories’ respective features as they might relate to the three test cases of delusion previously specified, i.e. erotomanic delusion, Capgras delusion and perceptual delusional bicephaly, with the discussion organised around the three foundational principles each theory makes use of (or rejects) regarding underlying neuropathology, (deep) irrationality, and abnormal experience.
The conclusion I draw from the analysis is that while, on the face of things, each of the established accounts of delusion exhibits strengths and weaknesses, on closer inspection no one theory stands out as being able to meet the first thesis question, not only because no particular theory is able to successfully meet the data set of clinical features re the three test cases of delusion, but because, when tested against these three cases of delusion, each of the theories encounters a knot of problems regarding each of the three categories of assumption to which they respectively appeal.

Finally, I would like to thank Professor Paul Noordhof and Dr Louise Richardson for their support and feedback on drafts of this chapter.

The Three Prevailing Theories of Delusion

The Empiricist One-Factor Theory

In chapter one, I introduced the empiricist view of delusion and explained it as being the hypothesis that in any case of delusion the key explanatory factor is by necessity a (clinically) abnormal experience of some kind, and that the fact of this abnormal experience provides the stimulus for subsequent cognitive processing to the end that a related (factual, propositional) belief, qua delusion, takes hold.

Two further points which were observed concerning the empiricist model are that: First, the nature of the abnormal experience is actual phenomenal consciousness which happens in virtue of actual causal changes in the neural machineries in the experiential division(s) of the living brain, and that this is the case whether the experience in question is a veridical perception, an illusion, or a hallucination. That is to say, according to the empiricist model experience in perception and in hallucination is common in kind or ‘commonkind’ (e.g. Macpherson 2013). And second, within the brain, the neural machineries of experience and the neural machineries of cognition are separate, i.e. each mental state is the product of a discrete set of neural correlates, with prior experience providing the essential evidences for subsequent (factual) beliefs. Although it is sometimes argued that cognition and other non-experiential states, such as moods, emotions etc., can ‘penetrate’, i.e. affect the nature (content, character) of a subject’s actual experiences (e.g. Siegel 2010, 2012).

The above ideas hold for any empiricist approach to delusion. However, if the account in question is an empiricist one-factor account, then there is the additional
stipulation that following the key abnormal experience all subsequent cognitive processing is fundamentally normal (Maher 1974, 1999, 2005). That is to say, other than anomalous experience, by necessity, no further mental deficit is awarded an explanatory role. So according to the empiricist one-factor thesis, any patient presenting with delusion is, all things being equal, deemed a fundamentally rational agent, insofar as his or her delusion is in actual fact a reasonable response to a highly unusual experiential event. The empiricist one-factor theorist may support this idea using cases of monothematic delusion, such as Capgras syndrome (e.g. Sullivan-Bissett and Noordhof 2015), in which by definition the delusion is a contained mental aberration within an otherwise normal belief set.

So if one were to plot the argument that the empiricist one-factor theory of delusion is making, it might look something like this:

P1  Experience is actual and common in kind.
P2  (Factual) Beliefs develop via a cognitive process of attempting to explain (actual, commonkind) experience.
P3  On the whole, normal perceptual (actual, commonkind) experience provides for true beliefs.
P4  In some cases, abnormal perceptual (actual, commonkind) experience can reasonably enough engender false beliefs.
P5  Delusion is a (factual) belief, but by definition it is false.
P6  Yet in cases of monothematic delusion at least, by definition, the delusion is a particular false belief in a global belief set of true beliefs.
P7  The falsity of the delusional belief has to start somewhere, and the causal root must be a particular abnormality of the brain.
P8  However, in the case of monothematic delusion at least, there is no explanatory warrant to posit an abnormality in the neural machineries of cognitive functioning.
C  Therefore, in the case of monothematic delusion at least, the explanation must be that an earlier abnormal (actual, commonkind) experience, the causal root of which is located in an abnormality of brain’s experiential machineries, is the key explanatory cause.
The first two premises (P1-P2) of this model reflect the empiricist proposals outlined above, and if we accept these premises then the next two premises (P3-P4) are basically common sense statements which follow on. The next part of the argument (P5 onwards) turns to the empiricist one-factor view of delusion. P5 bluntly states that delusional belief is a belief like any other belief. And given this, P6 observes that in cases of monothematic delusion the delusional belief stands out as the sole significant cognitive anomaly affecting an otherwise cognitively normal subject. Next a superficial reading of P7 is that it is observing that every effect must have an equivalent cause. Yet more than this P7 denotes the idea that the empiricist has a requirement to locate said cause in the structures and functions of a subject’s living brain, and specifically in a significant neuropathology because (a) according to the empiricist model the structures and functions of the living brain are where experience and cognition come about; and (b) without the stipulation of a significant neuropathological root cause the one-factor model over-predicts, for the reason that, given the key idea that cognitive processing is fundamentally normal (P8), there is no clear curb preventing all cases of abnormal perceptual experience (cf. P4) from developing into delusion. Next P8 provides warrant for the idea that cognition is fundamentally normal by observing that although it might seem obvious to posit a deep disorder in the cognitive workings of the brain as the explanatory root of the problem, the anomalous effects of this sort of abnormality would not be contained in the way that a monothematic delusion is a contained cognitive disorder. Therefore, the argument concludes (C), the crucial explanatory abnormality in cases of monothematic delusion at least must be deeply abnormal experience alone.

The argument seems to be valid. Nevertheless, some concerns immediately leap out. Most obviously, the whole proposal rests on the opening empiricist premises, but if one doesn’t support the empiricist suppositions, then the remaining case is doubtful to say the least. In response, it could be argued that P1 and P2 are merely the one-factor theorist’s foundational premises and as every argument has to open somehow, so these premises should be allowed to stand. And for sure, this is good practice. Nevertheless, what can be done is to draw attention to the problems of maintaining an empiricist model in re the cases of delusion which I use as test cases. And I do this presently.

In addition, another area of concern is P7. First, there is the idea that a particular delusion must have (and only have) a specific parallel cause in the structures and workings of the living brain. This is an empirical assertion which I am unwilling to
simply take as given. Although whether it is true or not is a question which I must leave for others who know more than me about causal relations in a biological system to decide. Yet a second concern with P7, and a point which can be logically unpacked, is the idea that the foundational cause of the anomaly must be rooted in neuropathology. And this is the first of the assumptions I address later in this chapter.

The Empiricist Two-Factor Theory

The empiricist two-factor account rests on the same foundational assumptions as the empiricist one-factor account regarding the nature and role of experience and belief. But then, thinking back to the empiricist one-factor argument plotted earlier, the two-factor theory breaks from one-factor theory somewhere around P5 – if P5 is read as the claim that a delusional belief is a belief like any other factual belief – for the reason that the two-factor theorist sees something different about delusional belief.

The difference the two-factor thesis discerns does not lay in (if you like) the fabric of the belief, since both accounts endorse the same empiricist principles when it comes to what belief is. Instead, the difference has something to do with the epistemic status of the delusional belief as compared to that of ordinary false beliefs, in so much as the falsity characterising a delusional belief is seen as being far deeper and far more resistant to correction than the falsity characterising the sort of commonplace anomaly proposed by P4.

For instance, under normal circumstances if one reasonably but mistakenly thinks that the stick which is half submerged in water is in fact bent, this belief would be seamlessly corrected should the stick be removed from the water and one sees that it is in fact straight and/or the principle of refraction is explained. However, this is markedly different to a case in which despite proofs and reasoning to contrary, Mr A unswervingly insists that his entire family has been replaced by near identical duplicates, or Mr P categorically maintains the conviction that he has in fact acquired a second head.

In view of this observation, the empiricist two-factor theorist suggests that certain constraints on belief which normally limit the content and/or the role of belief are deeply compromised in cases of delusion e.g. (e.g. Coltheart et al. 2011 pp. 283-6). What causes the compromise? Not experience, because as the stick example demonstrates mere anomalous experience is insufficient to explain the intractability of
delusional belief. Hence the two-factor theorist proposes the necessary fact of a second neuropathology, in this case affecting the structures and workings of cognitive processing, and which has an effect of undermining the system’s normal cognitive constraints (ibid.).

This is to say, the empiricist two-factor thesis is the view that in any case of delusion there must be two concurrent explanatory factors. The first factor, in agreement with the fundamental empiricist view, is abnormal experience. The second factor is deeply abnormal cognitive processing, so that individually either factor is insufficient to result in a case of delusion, but jointly they are the necessary and sufficient causes for a delusion to occur. And each factor is at root explained by the fact of a discrete pathology of the brain (ibid.).

So if one were to plot the argument that the empiricist two-factor theory of delusion makes, the initial four premises would be the same as those of the one-factor theorist’s argument, but then things would change around P5 when the argument moves to thinking about delusion, so that the remainder of it might look something like this:

P5  Delusion is a (factual) belief, but by definition it is false.
P6  Nevertheless, the false beliefs characteristic of delusion are different from normal false beliefs characteristic of those implied by P4, in that the falseness of a delusional belief is deeper and more resistant to change.
P7  This difference in the case of delusion is suggestive of a deep-rooted problem concerning a lack of cognitive constraint.
P8  And the problem has to start somewhere, namely somewhere in the structures and workings of the cognitive brain.
P9  However, on its own abnormal cognitive processing is insufficient to explain delusion, namely the content of the delusional belief has to be explained.
P10 And we know from P4 that abnormal perceptual (actual, commonkind) experience can engender false beliefs, but [as explained above] given the requirements of the empiricist approach to delusion, in cases of delusion, the root of the experiential abnormality must be located in an underlying abnormality of the brain.

Therefore, in any case of delusion there are two explanatory factors. The first, an abnormality of experience, is primary, and the causal root of which is a pathology of the experiential brain. The second, an abnormality of the cognitive
system is subsequent, and the causal root of which is a pathology of the
cognitive brain. And the factors are individually insufficient but jointly necessary
and sufficient to explain any case of clinical delusion.

The revision at P6 might seem to be a reasonable enough amendment. However, again,
there are problems which start to jump out. For instance, any general drawback
concerning an appeal to neuropathology seems to compound in line with the idea that
in any case of delusion the subject must necessarily be afflicted by not one but two
discrete and significant pathologies of the brain (P8, P10, C). Plus there is a feeling of
concern regarding the empiricist two-factor theory's appeal to deep irrationality for at
least two reasons.

First, as was noted previously, while one may wish to be cautious about making
conjectures concerning specific causal pathways in a complex biological system,
somewhat awkwardly for P7, it doesn’t seem entirely unreasonable to think that a deep
(upstream) deficit affecting the structures and workings of cognitive processing perhaps
ought to effect a wider array of (downstream) oddities than a specific conviction that
that person has been replaced by an imposter or this (second head) is my late wife’s
gynaecologist in generally normal-enough cognitive set.

And second, concerning P9 and P10, there is a nagging concern about the
division of labour between the two proposed explanatory factors as regards which if
either is doing the bulk of the explanatory work. It is a concern which becomes more
apparent in relation to delusions which characteristically have zero experiential element,
such as erotomanic delusion. And I return to this point when I consider the second
assumption of deep irrationality soon.

The (Relationist) Rationalist Theory

I take the prevailing rationalist theory of delusion to be the approach developed by
John Campbell (2001). And, as was outlined in chapter one, Campbell’s way of thinking
about delusion rests on the relationist view of perceptual experience to which he
subscribes (e.g. Campbell 2002).

To recap. As regards perceptual experience the relationist holds that whenever a
subject experiences a item in the observable world he or she does so in virtue of seeing,
hearing, feeling etc. the actual item, directly, as opposed to perceiving it indirectly via a
mental representation in the brain. In principle, the nuts and bolts mechanics and
cognitive workings of the brain are said to focus on the observable item in question
while pulling together the various facts of it into a coherent cohesive whole, so that, for
instance, when one sees the tree, the actual phenomenal contents of one’s experience is
literally supplied by the shapes, textures, colours etc. of the tree itself (e.g. Campbell

However, this raises a question about hallucination, and especially about (very)
strong perceptual hallucinations. Consider. On the one hand, as per the relationist
thesis, if we take it that by definition hallucinations are experiences of seemingly
observable stimuli which occur in the absence of a relevant observable stimulus, then
there is no relevant object and properties to supply the content of said experience, so
there can be nothing-it-is-like for one to hallucinate a tree (ibid.). On the other hand, if we
turn to the characteristic features of psychosis, then hallucinations are defined as being
perception-type experiences which are ‘vivid and clear, with all the full force and impact
of normal perceptions, and not under voluntary control’ (APA 2013 p. 87), and
moreover such hallucinations are a key diagnostic feature of certain psychoses, such as
the delusion of bicephaly undergone by Mr P. Thus it seems not unreasonable to call for
an explanation about what is going on.

And a typical way in which the relationist answers this query is by maintaining
the idea that by necessity there is nothing-it-is-like to hallucinate, but that in such cases the
brain effectively tricks the subject into thinking that, for example, they see a tree (e.g.
Macpherson 2013). In other words, there is nothing-it-is-like but only ever something-it-
seems-like to hallucinate a tree or a second-head. The relationist’s explanation is
essentially that in a brain which is equipped for perception via its machineries and
and cognitive capacities – and which cognitive capacities, according to Campbell and pre-
empting his theory of delusion, comprises two classes of belief, i.e. regular factual beliefs
and prior to these a background set of more intractable Wittgenstinian type framework
propositions (Campbell 2001) – in a certain situation applies its perceptual strategies in
an abnormal way and as a result misleads the subject into thinking that they are
experiencing x when in truth not, they are in fact experiencing nothing at all it only
seems to them that they are experiencing x, and moreover this can potentially happen with
all the vividness clarity force and impact of normal perceptions, so that in the absence
of reality testing it would be impossible for the subject to tell a normal perception and a
corresponding hallucination apart (ibid.).
So, if we plot these ideas, the background relationist part of Campbell’s rationalist account of delusion might look something like this:

First the relationist foundation:

P1 Perception and hallucination are experiential-type states.
P2 The good case of perception entails:
   a. observable objects plus cognitive processing;
   b. actual phenomenal experience.
P3 The bad case of hallucination entails:
   a. cognitive processing alone;
   b. zero experience.
C1 Therefore there is a disjunction between perception and hallucination, i.e. experience in the two states is not common in kind.

Then taking a closer look at the role of cognition and hallucination:

P4 In perception and hallucination cognitive processing plays a central role.
P5 The cognitive processing in question entails:
   a. focussing by the brain;
   b. the mapping and binding of information;
   c. factual beliefs;
   d. framework propositions.
P6 In perception and hallucination, a., b. c. d. can function in an identical way and so can realise what seems to be an identical experience, although:
   a. in perception, experience is actual in virtue of actual perceptual content provided by the actual item;
   b. in perception-like hallucination, there is no experience, only seeming experience, because there is no item supplying actual perceptual content.
C2 Therefore in perception and in an identical perception-like hallucination the subjective experience is fundamentally dissimilar, whereas cognitive processing is common in kind.

So much for the background thesis concerning perceptual type experiences, but what about delusion? In chapter one, I outlined Campbell’s ideas about delusion as regards
the background assumptions or ‘framework propositions’ held by the subject (or the subject’s brain), such as: *The world has existed for quite a long time*, as the set of these foundational assumptions changes to include an anomalous assumption, such as: *My neighbour has been replaced by an imposter* (Campbell 2001 pp. 96-7). And where the causal root of abnormality is a deep pathology of the brain (pp. 97-8). And where fact of the abnormal foundational assumption skews subsequent reasoning (pp. 96-8), and maybe as a downstream consequence fine grained functional and coarser grained public behaviours, such as the body’s autonomic responses and the subject’s linguistic outputs (pp. 97-8). And where any or all of these changes might in turn effect further changes to the subject’s experiences, for instance re atypical feelings of affect, but where any anomalous experience is only ever and at best a tangential downstream consequence of the delusion (pp. 98-9), the sole significant explanatory cause of which is always located in anomalous cognition, i.e. deep in the subject’s foundational assumptions, alone. In providing this explanation Campbell focusses largely (but not solely) on the case of Capgras delusion. However, as was also observed in chapter one of this thesis, different delusions are characterised by dissimilar characteristics. And so to provide an account of delusion *per se*, Campbell’s (relationist) rationalist approach at least needs to address the different clinical features outlined here as the data set, and which I have exemplified by the three cases of delusions as point.

So mapping the (relationist) rationalist approach to delusion, it might look as follows:

**P7** On the whole, a normally functioning system gives rise to accurate experiences and true beliefs.

**P8** Delusion is by definition a false belief, with clinical cases involving:

   a. zero experiential element;
   b. a mild experiential element;
   c. a (very) strong experiential element.

**C3** Thus to tend a cohesive account, a theory of delusion needs to address cases of types 8a, 8b and 8c.

**P9** (From P3 and P6b) In cases of delusion involving hallucination(s), the relationist cannot appeal to actual experience.

**C4** Therefore for the relationist, the crucial anomaly within the system which
underlies all cases of delusion must be located in the cognitive (belief formation) system alone – which founds a (relationist) rationalist account of delusion.

Note. This is not to suggest that hallucination is the only or even a primary reason for a relationist to adopt a rationalist approach to delusion, it only shows that it is a sound reason for the relationist to do so.

P10 In identifying the crucial cognitive anomaly, from a relationist standpoint and given the workings of the cognitive system (see P5), the locus of delusion must be somewhere in:
   a. focussing by the brain;
   b. the mapping and binding of information;
   c. factual beliefs;
   d. framework propositions.

P11 Only framework propositions have the deep epistemic resistance to change which is characteristic of delusions.

C5 Therefore for the (relationist) rationalist the crucial cognitive anomaly which underlies all cases of delusion is inside a subject’s framework propositions.

So if a corruption in framework propositions, i.e. in the deep underlying assumptions a subject (brain) holds, is the locus of delusion, then as regards the development of a clinical case of delusion, the (relationist) rationalist argument for delusion can be finished thus:

P12 The crucial upstream anomaly concerning framework propositions progressively effects various abnormal and significant consequences with regard to downstream functions, behaviours and (seeming) experiences, e.g. the subject may avow: ‘This person [his wife] is not my wife, it is an imposter.’

C6 Thus the specific clinical features characteristic of a particular delusion or delusional syndrome manifest.

C7 And so should abnormal experiences feature, they are always an effect of the delusion, never its cause.
This completes the outline of Campbell’s (relationist) rationalist approach to delusion. And again there are early concerns which jump out. For instance, the assumed fact of a significant neuropathology is again a factor. In this case, it is said to debase the structures and workings of the cognitive system, and so supply the explanatory root of the anomalous cognitions at point. However, even if one were to grant Campbell’s point that a blind pathology or ‘organic malfunction’ could effect a specific anomalous foundational proposition (Campbell 2001 pp. 97-8), as I argue in a moment this is really not an assumption which can or should be generally made.

Then there is the problem of positing deep irrationality (C4 onwards). Not least because, even if one can accept the idea of a deficit in a subject’s core foundational assumptions to the effect that, say, My wife has been replaced by an imposter, it still seems reasonable to think that this sort of deep cognitive corruption is going to have widespread effects, or at least effects which are more widespread than the abnormalities characteristic of a monothematic delusional disorder, such as Capgras syndrome – as Campbell observes concerning the supposed effects of gaining a significant piece of new information, ‘there would be endless reverberations of this revision all through my system of beliefs’ (p 98).

Finally, Campbell’s (relationist) rationalist account highlights potential problems with the third assumption of abnormal experience – though as I argue presently, this is an area of concern which is not limited to Campbell’s account alone. The decisive concern for the (relationist) rationalist view of delusion is the problem of whether or not a relationist approach to hallucination (P1-C1, P4-C2) can plausibly account for the kind of very strong hallucinations which are a defining characteristic of cases such as that of Mr P, whose hallucinations of a second head are so (seemingly) perceptually vivid and immediate that in his despair Mr P considers removing the head with an axe before attempting to shoot it dead.

This ends my mapping of three prevailing approaches to delusion. Yet as was noted in each case, it is reasonable to suggest that the respective theories might face problems re the key assumptions they variously make concerning neuropathology, and/or rationally, and/or experience. Thus in the second part of the chapter, I examine each of these assumptions in turn.
Appeal to Neuropathology

Each of the prevailing theories of delusion makes appeal to the fact of a clinically significant prior abnormality the effects of which realise the explanatory bases of any given case of delusion. There is also agreement that the prior abnormality in question is a disorder of the living body. That is to say, all delusion is held to be ‘biologically based’ (Maher 1974 p. 110) involving an ‘impairment of function’ (p. 99); or underlying ‘organic medical conditions’, ‘organic brain damage’ etc. (Coltheart et al. 2011); or ‘organic malfunction’ (Campbell 2001 pp. 97-8), and specifically one which undermines the integrity of the living brain. Namely, a neuropathology (e.g. Coltheart et al. 2011 pp. 285-6). And all the theories hold that the abnormality or pathology is clinically significant, which is to say it is not a something and nothing, but by explanatory necessity it is a something of substance, i.e. a case of biological abnormality which brings about actual and problematic disturbances in the way the living body works, and the effects of these disturbances supply the specific explanatory reason(s) for the clinical problem at point. Namely, the anomalous experiences and/or deep irrationality which the delusional patient is variously said to undergo.

Why make the assumption of a necessary underlying neuropathology? One, it is a direction in which explanations have moved in recent times. For instance, as Maher writing in 1974 observes, the ‘biologically-based hypothesis accepting the reality of perceptual anomalies only gains plausibility in the light of newer knowledge about the neurological determinants of consciousness’ (Maher 1974 p. 110). Two, if the requisites a theory depends on are identified as properties of the biological brain, then the explanatory options for identifying the root cause of the requisites are limited to factors inside and/or affecting the integrity of a normal brain and which by necessity effect a sufficient disturbance to addle the biological underpinnings on which the requisites – which in this case means experience and/or cognition – are said to depend.

Nevertheless, a blanket appeal to significant neuropathology as the necessary causal root of the explanatory requisites of delusion is not a simple assumption, and from whichever angle the claim is approached it generates not insubstantial snags for each of the prevailing accounts.

For instance, if what constitutes the significant neuropathology is delineated in a strong sense, which is to say, as the sort of problem which shows up in brain imaging data as a (relatively) unambiguous anomaly which, unless something is done about it in
terms of surgical and/or medical intervention, would not tend to go away – for example, lesion(s) caused by traumatic injury, haemorrhage or tumour or, perhaps, a significant change in brain tissue caused by, say, degenerative disease – then, in the case of a patient suffering from delusion in whom investigations identified this sort of neurological change, it is likely that some sort of relation could be drawn between the brain damage and the delusion at point. Consider, for instance, the case of erotomanic delusion which happens in conjunction with significant and hitherto undiagnosed vascular lesion of the brain (Farnham et al. 1997).

However, even in cases such as this in which, on the one hand, the identification of brain pathology and, on the other hand, the diagnosis of delusion are wholly unambiguous, the nature of the relation between these factors is not always (if ever) clear-cut. Not least because further conditions, such as the subject’s wider environmental, social and/or biographical circumstances, are impossible to factor out of the reckoning. Indeed, in all of the case studies outlined in chapter one, whatever else may (or may not) be going on with the subjects, the wider special or difficult circumstances with which they variously contend is a notable feature of every case. I say more about this in a moment.

Furthermore, even if we allow that a significant correlation does in fact exist between cases of strong neuropathology and cases of delusion, and further allow that said neuropathology is sufficient to effect said delusions, a question remains as to whether strong neuropathology could be a necessary explanatory factor with regard to all cases of delusion especially as evidence suggests that this possibly (indeed, probably) is not the case.

For instance, in very many cases of persons who have received a diagnosis of delusion, either given the person’s symptomology further empirical investigations, such as brain imaging, are not warranted and so do not happen, or if for whatever reason further investigations do happen they do not reveal the presence of any significant changes in the brain. Consider, Capgras delusion as outlined in chapter one. Although co-occurrences of various sorts have been observed in a good number of cases, this is certainly not true for all cases of Capgras delusion tested. And also from chapter one, consider erotomanic delusion. It is a relatively common disorder in which wider special or difficult cultural, social or biographical circumstances are generally recognised to strongly correlate, but as regards which delusion further investigations very rarely detect any notable changes to the subject’s brain.
Nonetheless, it might argued that the reason that neurological changes are not detected in subjects suffering from delusion is because (some of) the changes in question are outside the remit of that which current brain-scanning technology can detect. Thus the right approach is to extrapolate from cases of delusion where co-occurrence is observed to all other cases of delusion and predict that a significant brain pathology is present and to then seek evidence of that pathology as best as one is able at the time (cf. Coltheart et al. 2011).

However, while it is surely the case that there are limitations to current investigative techniques, one must bear in mind that what is actually being proposed here is an instance of inductive reasoning, which by its very nature selects one likelihood from a whole set of likelihoods. It is not certainty. And given the evidence as it stands it is arguably not even the most likely option on the table. Indeed, a wholly plausible alternative hypothesis would be that in the cases of the very many subjects who suffer from delusion but for whom clinical investigation does not indicate the presence of significant neuropathology, the reason for this result is that said subjects simply do not have any such damage to their brains. And again, this is not to suggest that imaging data or the interpretation of imaging data is infallible, because such a claim would be patently inaccurate. Nevertheless, across other clinical fields the exact same sort of empirical data is relied upon to disclose that which needs to be disclosed reliably enough, so one might wonder what is so special about delusion that it manages to sidestep the investigative techniques that, say, trauma specialists, oncologists and neurosurgeons rely on every single day.

Moreover if one steps back and considers the epidemiological data it is arguable that one wouldn’t (oughtn’t) actually want there to be a necessary correlation between strong neuropathology and the occurrence of psychotic features, such as delusion. The reason for this is that epidemiological studies indicate that hallucinations and delusions, both transient and enduring, are prevalent throughout the general (healthy) population (e.g. Johns and van Os 2002; Verdoux and van Os 2002; Rossler et al. 2007). So if we take the case of all subjects for whom a current diagnosis of delusion or delusional syndrome has been made; plus those for whom such a diagnosis has been made in the past but who are now recovering or in remission –remember, the strong pathological damage in question isn’t a kind which simply tends to go away; plus those who suffer from psychosis or psychotic disorder but whom, for one reason or another, stay off of the clinical radar; plus a substantial number of normal healthy people who variously
experience delusions, hallucinations etc. but for whom a diagnosis of psychosis or psychotic disorder would not now or ever be made; and if we factor in the point that in a substantial number of cases of those diagnosed with delusion and who have undergone neurological investigation in relation to this diagnosis, the results of the investigations are that no abnormality is detected; then at this point would (ought) one really want to make the a priori assumption that all of these people are necessarily suffering from one (apropos of the empiricist one-factor and rationalist relationist accounts) or two (apropos of the empiricist two-factor account) significant instances of the dreadful and typically debilitating sorts of brain injury/disease under discussion? My own view is that, no, I balk at the idea of doing so – though, of course, each to his or her own.

All the same, perhaps it might be argued that, okay, even if it’s implausible to propose that a significant structural abnormality of the sort described is a necessary underlying factor of delusion in every single case, the situation could still be that the pathology in question might be less obvious and perhaps more pervasive, for instance something closer to the early-stage, sub-clinical (premorbid) neurological changes associated with neurological disease. My response to this would be that, personally, I don’t find this suggestion plausible either, plus premorbid changes hardly seem to be instances of the theoretically required significant brain damages at point.

Still, maybe the point could be pressed. For instance, it might be argued that, given the problem of defining the (proposed) necessary abnormality of brain in terms of strong neuropathology, the abnormality should instead be defined in weaker terms (cf. Ellis and de Pauw 1994), say with reference to abnormal functional, electrical and chemical changes of a sort which may be invisible to some brain monitoring/imaging technologies, but accessible to other kinds of analysis, such as histological investigations? Plus unlike strong neuropathology, qua lesion or significant degeneration of brain tissue, which once present are unlikely simply to go away, instances of weaker neuropathology could be more fluid having transient effects which might come and go in response to various environmental and bodily stimuli or which could be fine-grained indicators of an enduring biological disorder, such as migraine, temporal-lobe epilepsy, or dementia.

An advantage of maintaining that weaker pathology underlays the explanatory factors of delusion is that maybe it goes some way to explaining the transient nature of some cases of delusion, not only in the context of clinical psychiatry, but across the demographic. For example, it suggests a way of addressing the similarity between the
acute symptoms of psychosis and the use of certain drugs, such as benzodiazepines, amphetamines, hallucinogens and cannabinoids (e.g. Corlett et al. 2009).

Furthermore, if features of psychosis, specifically hallucination and delusion, can now be associated, via weak neuropathology, with wider environmental circumstances and physiological changes, then it opens the way for explanations in terms of, say, foodstuffs, drugs and other chemicals; various activities and lassitudes; all manner of stressors, pleasures; affective states, anxiety, arousal, moods, elation, distress and so on, which, on the face of things, sits better with the epidemiological data since we are no longer having to presuppose that a significant number of the population suffers from either one or two clinically significant occurrences of brain damage.

However, a downside of positing weak neuropathology is that it becomes harder to account for long-term (tantamount permanent) cases of delusion and delusional syndrome – unless we, once again, return to the idea that said delusions must be an (early) indicator of progressive disease, such as dementia, but which in turn doesn’t explain why the content of delusions can remain fixed for long periods of time, even decades, neither varying, intensifying nor compounding. (See the outlines of erotomanic delusion and Capgras delusion in chapter one.) And again while there is evidence to suggest that there could be some sort of correlation between weak neuropathological changes of a sort associated with, for instance, migraine or uncontrolled diabetes, and Capgras delusion (Ellis and de Pauw 1994), there are other delusions, such as erotomanic delusion, in which pathologies of even a weak kind are not indicated.

What is more, in making appeal to weak neuropathology, we are allocating an explanatory role to fine-grained functional, electrical and chemical changes which are not only associated with disorders and diseases of the brain, such as migraine and epilepsy, but also with the physiological functions and processes which occur in normal responses to environmental factors, such as stressors and pleasures, and in connection with affective states, emotions moods etc. That is to say, the line between abnormal and normal has vanished, and we are no longer appealing to neuropathology as such, but to neurological change, which may or may not be within a range of normal.

In other words, while the electrical or vascular changes associated with migraine may well correspond in kind to the sort of fine-grained vicissitudes which might underwrite delusion, so too could the non-pathological changes associated with normal responses to certain environmental stimuli. And sure, the internal versus external ontology of the two explanations may, on the face of it, look dissimilar. And the causal
pathways involved may, on the face of it, look different – for instance in a case where the brain has to deal with abnormal physiology qua an intrinsic bodily condition the pathway might be shorter. But if, in both cases, the effect is the same, namely the subject undergoes psychosis, so either the weak neuropathology associated with, say, migraine, or the normal neurophysiology associated with, say, difficult or special environmental circumstances, would be a legitimate cause of the psychiatric disorder. And, indeed, while, on the one hand, there may well be a relation between some cases of Capgras delusion and certain bodily disorders such as migraine and diabetes; on the other hand, it is widely acknowledged that there is a marked relation between cases of erotomanic delusion and the subject’s wider environmental, cultural, social, and interpersonal circumstances, as was outlined in some detail in chapter one.

However, now the problem compounds, because if weak neuropathology qua neurological change is held to underwrite cases of delusion and delusional disorder, then one may well expect the incidence of said delusions to be even more prevalent than the epidemiological data suggests. For instance, if we take the case of all subjects for whom a current diagnosis of delusion or delusional syndrome has been made; plus those for whom such a diagnosis has been made in the past but who are now in recovering or in remission; plus those who suffer from psychosis or psychotic disorder but who, for one reason or another, stay off the clinical radar; and add to this the substantial number of normal healthy people for whom a diagnosis of psychosis or psychotic disorder would not now or ever be made, but who suffer from migraine, temporal lobe epilepsy, diabetes etc.; plus everyone who at some time or another has to deal with tricky environmental, social, interpersonal etc. circumstances, and if we then predict that all these people will suffer either transient or long-term delusions or delusional syndrome, well, that amounts to a lot of delusions. So while fine-grained functional, chemical or electrical changes may perhaps be a necessary factor as regards cases of delusion, it is fair to say that they are probably not a sufficient cause.

**Appeal to (Deep) Irrationality**

The second assumption applied by the prevailing theories of delusion concerns the factor of rationality, or its inverse irrationality. And regarding this feature, the empiricist one-factor theory stands alone in holding that ceteris paribus the subject who suffers from delusion is by necessity a fundamentally rational agent. Whereas the empiricist two-
factor theory and Campbell’s (relationist) rationalist account maintain that by necessity irrationality is a second or the sole explanatory cause.

To clarify, the irrationality in question does not mean the sort of commonplace miscalculations of reasoning which everyone makes from time to time, maybe because of a failure to think things through carefully enough, or possibly because one is in a stressful or distracting situation, or perhaps because one isn’t in possession of or doesn’t sufficiently understand all of the pertinent facts. For instance, as with the (fallacious) train of reasoning: *All cows eat grass. Daisy eats grass. So Daisy must be a cow.* Therefore when the empiricist one-factor theory claims that the delusional subject is fundamentally rational, this in no way suggests that said subject is a perfect reasoner who is immune from making normal enough (if sometimes seemingly odd) miscalculations of thinking, it just means that there is nothing intrinsically amiss with the way in which the subject processes these thoughts.

In contrast, the empiricist two-factor theorist and Campbell’s (relationist) rationalist thesis both submit that in any case of delusion there must, by necessity, be a significant abnormality affecting the integrity of the biological brain, which (abnormality) addles the way the brain’s cognitive machineries function and so leads to deep abnormalities of thought. For a non-psychiatric example of this sort of deep abnormality and the sort of glitches of reasoning which can be generated, consider Oliver Sacks’ famous account of the man who mistook his wife for a hat (Sacks 1985 [2011]). For instance, when Sacks asks the subject, who he refers to as Dr P, to identify a glove, in some ways Dr P’s descriptions of the object he perceives exhibit a kind of uber rationality, even so the structural damage affecting the brain’s cognitive capacities specifically scuppers Dr P’s capacity to name (identify, classify) the object at point.

> ‘What is this?’ I asked, holding up a glove.

> ‘May I examine it?’ he asked, and, taking it from me, he proceeded to examine it.

> ‘A continuous surface,’ he announced at last, ‘infolded in on itself. It appears to have’ – he hesitated – ‘five outpouchings, if this is the word.’

> ‘Yes,’ I said cautiously. ‘You have given me a description. Now tell me what it is.’

> ‘A container of some sort?’

> ‘Yes,’ I said, ‘and what would it contain?’
‘It would contain its contents!’ said Dr P with a laugh. ‘There are many possibilities. It could be a change-purse for example, for coins of five sizes. It could be …’ (Sacks 1985 [2011] p. 15.)

Thus the hypothesis being proposed by the empiricist two-factor theorist and (relationist) rationalist is that in psychiatric cases a specific biological problem deep inside neural workings of cognitive system effects downstream abnormalities of reasoning and performance, and which abnormalities realise or group together as the symptomologies of psychosis and psychotic disorder, and specifically, as cases of delusion and delusional disorder, such as occurs in Capgras syndrome, in which the key symptom is that the subject presents with the deep-seated conviction that: This person [for instance, the wife] before me now, is not in fact my wife, but is an imposter [of some sort].

And, on the plus side, it might seem that on the face of things it is plausible enough to suggest that a problem deep inside the workings of the reasoning processes would result in a deeply anomalous conviction of some sort. After all, by definition delusion presents as a cognitive incongruity. Thus perhaps a cognitive deficit should logically be awarded an explanatory role.

All the same, it is fair to say that problems soon emerge on account of positing a necessary role for (deep) irrationality. And in the remainder of this subsection I consider three possible concerns, each of which somehow relates to downstream effects which occur (or fail to occur) as a result of the proposed factor of an upstream anomaly of the cognitive system. The nub of each problem pertains more or less equally to the empiricist two-factor account and the (rationalist) relationist account, although given the theoretical differences between the two theses so the problem perhaps develops along a somewhat different pathway in either case. The three problems I identify are:

1. The problem of positing a general biological failure as the foundational cause of delusion and delusional disorder, particularly as regards cases involving monothematic delusional content.
2. The problem of specifying and/or constraining the significant abnormal cognitive feature.
3. The problem that the reasoning processes of subjects with delusion, and specifically with a monothematic delusional syndrome, fall within a remit of normal.
Further to these functional and/or performance concerns which I argue follow from a general appeal to the necessary role of deep irrationality, I observe another, fourth, concern. This time the issue applies to the (relationist) rationalist account alone, and it rests on the appeal made by Campbell to the notion of ‘framework propositions’. Essentially the concern is that as it stands the notion of ‘framework propositions’ is unclear and that when pushed it finds itself squeezed between horns of a significant dilemma. However, given the nature of the (relationist) rationalist thesis, so the issue only really gains traction in conjunction with the assumptions which are made by the (relationist) rationalist concerning abnormal experience. Therefore I defer discussion of this concern to the following subsection of the thesis.

1. The problem of positing general biological failure as the foundational cause of delusion and delusional disorder, particularly as regards cases involving monothematic delusional content.

Both the empiricist two-factor thesis and the (relationist) rationalist thesis posit the fact of a significant neuropathology affecting the regions of the brain which regulate cognitive functioning as the necessary and sufficient causal foundation of the (deep) irrationality which they maintain is a significant factor in all cases of delusion and delusional disorder. That is not to say that in every single case of delusion precisely the same type of neuropathology is held to be present. It is simply that both the two-factor theorist and the (relationist) rationalist uphold the idea that in any given case of delusion or delusional disorder the actual occurrence of one-or-another substantial neurological abnormality has the functional effect of throwing the workings of the cognitive system appreciably and significantly out of norm.

It is fair to say that for both sets of theorist the precise nature of the proposed biological abnormality remains something of a mystery, and which point is amply illustrated in discussions of various sources (e.g. Ellis and Young 1990; Ellis and de Pauw 1994; Coltheart et al. 2011).

Nevertheless, each approach is clear as to the nature and role of the resulting cognitive abnormality. Specifically, for the empiricist two-factor theorist, the ‘organic problem’ results in a general difficulty relating to the extraction of particular hypotheses from the set of all possible hypotheses, which therein tolerates the selection of hypotheses and with it conclusions which are improbable, bizarre or impossible (e.g. Coltheart et al. 2011 pp. 283-6), while, for the (relationist) rationalist, the ‘organic malfunction’ realises a foundational cognitive abnormality re some framework
proposition (i.e. foundational belief) against which the subsequent testing of other 'factual' beliefs (i.e. hypotheses) takes place (Campbell 2001 pp. 96-8).

Nevertheless, in either case, a problem with this sort of proposal is that, as the above mentioned case of Dr P illustrates, and as Sullivan-Bissett and Noordhof argue (Sullivan-Bissett and Noordhof 2015 p. 4), one would expect that a subject suffering from a pathology involving a region of the brain associated with essential cognitive functioning would present with a more general cognitive problem than is typically associated with cases of delusion and especially with a monothematic delusional syndrome, which by definition entails either a single delusional belief or a small set of delusional beliefs which relate to a single topic.

In response to this observation, the empiricist two-factor theorist might argue that the specific content of the delusion is determined by the phenomenal content of the specific experiential anomaly which the damaged cognitive system simply works to comprehend, conceptualise or ‘make sense of.’ However, such a response is a non sequitur. Since either the cognitive system isn’t damaged or it is damaged. And if it isn’t damaged, then the claim that it is ‘simply working to make sense of’ inputted data might (in theory) work. However, if it is damaged then there is no ‘simply’ about it. Indeed, given the manifold of experiential data with which any subject’s perceptual system is bombarded at any moment, and given the ongoing train of hypotheses which the empiricist maintains must be generated to make sense of this onslaught of data, it is implausible to say that the effect of the cognitive anomaly is bounded by the occurrence of an abnormal datum. Indeed, if the cognitive system is as deeply cockeyed as the two-factor account proposes, then it is entirely feasible that the contents of the abnormal experience be inferred correctly, while the contents of a normal experience be inferred incorrectly. In other words, the sort of damage to the cognitive system posited by the empiricist two-factor account generates a situation in which anomalous cognitive contents ought to crop-up all over the place.

The relationist, in turn, might simply say that this problem doesn’t apply to them since the relationist position is that the specific organic malfunction effects a specific cognitive anomaly re a specific framework proposition which specific abnormal framework proposition is precisely the specific delusional content with which the subject must contend. Yet this line of reasoning falls absolutely flat for, as Campbell himself observes while discussing his own theory, a single organic malfunction effecting a specific change in a foundational belief would have ‘endless reverberations ... all
through my system of beliefs’, and ‘in fact, it is hard to see a difficulty of principle here’ (Campbell 2001 pp. 97-8). And one has to wholly agree with Campbell on this particular point, for it seems evident that as per his (relationist) rationalist theory of delusion, a foundational error would indeed have ‘endless reverberations’ throughout a subject’s entire cognitive field. In other words, once again anomalous cognitive contents would crop up far and wide.

2. The problem of specifying and/or constraining the significant abnormal cognitive factor.

If for the sake of argument one takes as given the (deep) irrationalists proposal that a blind biological pathology affecting the area(s) of the brain responsible for cognitive function is the necessary and (perhaps) sufficient explanatory basis of any given case of delusion and delusional disorder, then the (deep) irrationalist might still be called on to specify the significant cognitive factor in question and/or explain how and why in cases of delusion, and especially in cases of monothematic delusion, the aberrant outputs of a pathologically damaged (i.e. deeply malfunctioning) cognitive system are limited to a particular and highly circumscribed set of anomalous beliefs while all other cognitive contents are on the face of it normal.

In answer to this question the empiricist two-factor theorist may elect to downplay the significance of the cognitive anomaly. For instance, the delusional subject might weigh data relating to experience over the effects of cognitive discrimination (Stone and Young 1997, in Sullivan-Bissett and Noordhof 2015 p.5). However, if the cognitive anomaly is significant enough to generate delusional cognitive content in response to anomalous experiences, then one would think that it would be significant enough to generate at least some anomalous cognitive contents in at least some cases of normal, or perhaps not normal but merely unusual, experiences too. However, not only does this not fit with the profile of monothematic delusion, but has been shown that delusional subjects tend overall to present as effectual enough (Bayesian) reasoners (Sullivan-Bissett and Noordhof 2015 p. 5). Whereas, if the cognitive anomaly is so weak that it only generates problems in cases of flagrant experiential abnormality, then one might query whether it warrants being credited with a significant explanatory role, which is to say, whether it adds anything of substance to the empiricist one-factor account.

On the other hand, the two-factor theorist might elect to meet the problem by positing a significant but highly specific neurological anomaly which explains how a deeply damaged cognitive system can reliably discriminate normal from abnormal
experiential data and, in the case of normal experiential data, generate belief contents appropriately, but, in the case of abnormal experiential data, generate belief contents inappropriately and so doing producing delusional content. However, in response to this proposal, one might question just how conceivable let alone coherent, such an explanation would be; and, if conceivable and coherent; just how clinically abnormal such a cognitive system would be; and, indeed, whether one needs to posit the second clinically anomalous factor at all.

Still another way in which the two-factor theorist might move to meet the problem would be to differentiate between the cognitive processes involved re the onset and re the development of the delusion. Yet if the two-factor theorist were to say that the anomalous function occurs at the onset of the delusion; which is to say, during the selection of the initial hypothesis from the set of all possible hypotheses in the functional roots of the cognitive system; it would still need to be explained why, in cases of normal experience, normal hypotheses are reliably generated, but in cases of abnormal experiences, an improbable/bizarre/impossible hypothesis is selected in the first place and, moreover, it is held. This is because, if the anomalous function occurs at the onset, and all subsequent cognitive functions connected with the holding or development of a belief are normal, then one would need to explain why the anomalous hypothesis isn’t de-selected somewhere downstream as part of a normal process of ongoing belief correction.

One possible candidate for this kind of explanation would be the idea that delusional subjects are generally disposed overweighting initial conclusions, and thereby fail to give sufficient weight to ongoing alternative hypotheses (Stone and Young 1977, in Sullivan-Bissett and Noordhof 2015, p. 5). However, even if we sidestep the fact that empirical doubt hangs over this proposal (Bentall and Young 1996, in Sullivan-Bissett and Noordhof 2015, p. 5), the idea still fails to account for the initial discrimination made deep within the neural machineries of the (malfunctioning) cognitive system re normal and abnormal experience, or how and why a system which mostly selects normal hypotheses will suddenly extract an implausible/bizarre/impossible hypothesis in response to abnormal experience alone.

Alternatively, the two-factor theorist may elect to shunt the cognitive anomaly downstream and say that while the extraction of initial hypotheses functions normally, the cognitive failure lies in the way that a subject’s beliefs are held and develop. In other words, experience may be normal or abnormal. If experience is normal, the initial
hypothesis selected will be normal, but if the experience is abnormal then the initial hypothesis selected will be ... normal (?) ... normally abnormal (?). If one thinks about it, it is somewhat unclear what it would be. Nevertheless, applying charity, we might say that the initial hypothesis would be odd but within the remit of understandable. However, during the subsequent clinically abnormal process of cognitive development, any quirks or oddities present in the initial hypothesis would compound (as opposed to being selected out), and, thus, the delusion develops.

This is, in fact, an approach favoured by Coltheart (Langdon and Coltheart 2000, in Sullivan-Bissett and Noordhof 2015, p. 5). However, the glaring problem with the proposal is that the hypotheses selected to account for normal experiences are subjected to the same processes of anomalous cognitive development as the hypotheses selected to account for abnormal experiences. So, unless each and every normal experience hypothesis is quite literally immaculate, flawless, perfect, one-hundred percent accurate straight out of the box (very-very unlikely), then one would expect even trivial glitches to gradually compound into significant cognitive anomalies. In other words, evidence of erratic, random and compounding irrationality would surface throughout the delusional subject’s ongoing behaviours, which would gradually devolve into something like a Brownian rationality (i.e. chaotic irrationality) over a period of time.

In comeback, those who support the idea that the necessary clinical cognitive anomaly posited by the two-factor theorist occurs downstream to the generation of a normal initial hypothesis might point out that an aberrant cognitive system could get things right as well as wrong. Hence, a degeneration into cognitive chaos is not inevitable. Nevertheless, successes would be happenchance. Indeed, the ongoing development of any cognitive inference would be the unreliable upshot of neurological, cognitive and/or circumstantial luck. And while this sort of pathological profile in some ways fits the profiles of certain degenerative neurological disorders – such as, certain kinds of dementia, in which moments of lucidity may occasionally, in the earlier stages of the disorder, at least, bubble-up as if from nowhere – it does not fit the pathological profile of monothematic delusion, or, indeed, psychiatric illness full stop.

Again, in reply, those who support the view might claim that the above pathological profile does, indeed, fit the pathological profile of some psychiatric cases, since some patients who initially present with monothematic delusion go on to develop the complex polythematic delusions indicative of major psychotic disorders, such as
schizophrenia. Yet while this is certainly true, it ought to be pointed out that such an escalation of symptoms is by no means inevitable, and that the monothematic delusions held by some other patients remain ‘fixed’ (indeed, in some cases remarkably fixed) for considerable periods of time. For instance, one patient was reported as holding a monothematic erotomanic delusion concerning the same object for over twenty-one years (Menzies et al. 1995 p. 530).

Moreover, even in the case of patients whose delusions do degenerate into polythematic delusions and who, subsequently, do receive diagnoses of a psychotic disorder such as schizophrenia, it would be grossly wrong to say that the cognitive functioning of any such patient is one of chaotic irrationality. Indeed, at times other than that of an acute phase of the condition, subjects who suffer from psychosis typically operate well within a remit of normal. Moreover, even in acute stages of endogenous psychosis – as opposed to the acute psychotic behaviours triggered by doses of psychotomimetic drugs (e.g. Corlett et al. 2009), which can seem highly chaotic to a naïve observer, but about which another story needs to be told – behaviours do not typically (if ever) deteriorate into absolute chaotic irrationality. It is more like, in such cases, the patient has fields of both rationality and irrationality, and even the fields of irrationality are not entirely chaotic, it is more like in these areas patients follow idiosyncratic pathways of reasoning and resultant behaviours, which although offbeat, abnormal or perhaps unsafe by public standards, are often internally cogent and highly predictable to knowledgeable observers, such as family and clinical staff.

So, in short, any attempt by the empiricist two-factor theorist to specify the abnormality of the cognitive system which is held to assume a necessary explanatory role re the occurrence of monothematic and, indeed, polythematic, delusions fails, because the proposed anomaly is either not significant enough and collapses back towards a one-factor account, or is too significant and generates too much in the way of anomalous output. Furthermore, the cognitive outputs (behaviours) of subjects who suffer from monothematic and polythematic delusions strongly suggest that no general cognitive anomaly, actually, exists.

The (relationist) rationalist, on the other hand, may well argue that their theory of delusion does in fact specify the significant cognitive factor in question. It is an abnormality concerning a framework proposition (or something which functions as a Wittgenstinian framework proposition is claimed to function) with the content of the
abnormal framework proposition being identical to the content of the delusional belief which is therein contained.

However, as was previously observed, the effect of the assumed ‘clinical’ abnormality would not be contained, since, as Campbell himself argues, the content of the abnormal foundational proposition would proliferate across subject’s cognitive map (Campbell 2001 pp. 97-8). And one reason why the content of the cognitive anomaly would proliferate is woven in to Campbell’s own account and concerns the principle of the epistemic feedback circle which develops. Essentially, at any time, ongoing facts (experiences, thoughts, etc) are checked against a map of background assumptions and as a result, when everything operates as it ought, an epistemically virtuous circle is instigated whereby background assumptions function as a reliable constraint on the ongoing flow of facts, deselecting those which are improbable or bizarre or simply unimportant, so that only facts with epistemic value to the system are retained, and, following the theory, it is these contents which potentially go on to affect the subject’s downstream experiences, actions, thoughts, expectations, ideas, feelings, moods, motivations etc.

However, concerning this process, the problem facing the (relationist) rationalist is that in delusion the subject’s cognitive framework is ex hypothesi abnormal. Hence, the feedback circle which plays out is corrupted, i.e. it is epistemically non-virtuous. In short, the presence of a significant background anomaly means that certain oddities of thought etc are not deselected but are retained, which realises odd downstream experiences, actions, thoughts etc, which odd outputs then feed back into the system, where they support and are supported by contents of the abnormal cognitive framework, thus affecting further downstream oddities, in an ongoing cycle the contents and outputs of which become progressively more deranged.

For example, Benny in his cognitive map holds the anomalous foundational assumption that: Rachel has been replaced by an imposter. Thus when Benny sees Rachel, the object he sees surely looks like Rachel. However, when he, or rather his system, tests the hypothesis that this is Rachel against the backdrop of foundational assumptions, the response received is that no, this cannot be Rachel, because Rachel has been replaced by an imposter, which conclusion the system takes at face value, which thus affects its downstream processing, which affects Benny’s ongoing experiences, actions, thoughts etc, and the contents of which experiences, actions, thoughts etc then feed back into the system for further testing, which in turn functions to authorise the content that this is
not Rachel, it is an imposter. Thus, from Benny’s, or rather the system’s, point of view, each cycle affords additional evidences for the notion that Rachel has been replaced by an imposter, which means that each cycle elevates the epistemic status of Benny’s abnormal belief, and so compounds the problem, and therein applies increasing pressures on Benny’s feelings, thoughts and behaviours, until the time comes when Benny publicly announces, ‘That is not Rachel, it is an imposter. How do I know? Just look! It’s obvious! I just know.’ (cf. Siegel 2012 pp. 201-3).

At this point, the relationist might interject that this description of Benny’s compounding predicament merely illustrates the compounding symptomology had by a subject who suffers from delusion or a delusional syndrome. However, there are two ways in which one might respond to this claim.

The first way is to point out (once again) that while in some cases delusional content does compound into deteriorating psychosis, this is certainly not always so. For instance, it is very likely that for many persons with delusion or delusion-type thoughts their condition never exacerbates to a point where a diagnosis of delusion is warranted and of those cases which are diagnosed, a patient’s symptoms can remain entirely stable in some cases for periods spanning many years.

The second way in which one might respond is to draw attention to the logic of the feedback cycle. If we accept the relationist thesis concerning the occurrence of an anomalous framework proposition, which in the case of Benny would be the foundational principle that: Rachel has been replaced by an imposter. And if we ask, what sort of ‘question’ (data, hypothesis) would the system test against this principle? Then it would be fair to say that a system would test not (only) the specific question: Has Rachel been replaced by an imposter? And not just the question: Is this Rachel before me now? But it would also doubtless test any and all propositions, memories, experiences, thoughts, ideas, expectations, moods, motivations etc. which have Rachel-related content against that particular foundational principle; and not only this, but there would also be all manner of merely tangentially Rachel-related content to test, such as, object content, person content, face content, woman content, wife content, sister content, voice content etc; and in addition there would also be imposter and imposter-related content to bounce off of the conceptual/normative backdrop, as well as countless other questions about goodness knows what contents which the brain might in that moment be moved to test. Then, consider that the outcomes of this first round of answers would also need to be enquired into as regards their importance and value as the system works
to fathom whether this content or that content meets with its reality and requirements (or not). Ditto any second round of answers, and third round of answers and so on. And all of this in a corrupted system which is not operating to constrain oddity but which is actively confirming and in so doing compounding oddity.

Hence, even if we were to accept the (relationist) rationalist proposal that in cases of delusion the significant abnormal cognitive factor is specifiable as a corrupted framework proposition, once we take into consideration the general hypothesis put forward by the account, it becomes evident that the (relationist) rationalist approach to delusion falters because it fails to afford a significant way of constraining the alleged foundational abnormality the abnormal contents of which would multiply exponentially across a subject’s cognitive map.

3. The problem that the reasoning processes of subjects with delusion, and specifically with a monothematic delusional syndrome, fall within a remit of normal.

One might think that, by definition, monothematic delusion precludes any explanation in terms of a necessary and clinically significant neuropathology of the cognitive system for the reason that even though the contents of a particular monothematic delusion may range from being merely implausible to clearly bizarre and impossible, the boundary of the delusion is, by definition, highly circumscribed in that it comprises a single delusional belief or a small set of delusional beliefs which relate to a single topic.

Furthermore, even in cases where the severity of the monothematic delusion warrants a diagnosis of psychotic disorder and further diagnostic criteria need to be considered, the further criteria actually serve to emphasise the bounded nature of the delusional content. For example, as part of the diagnostic criteria of delusional disorder, DSM-5 specifies that ‘Apart from the impact of the delusion(s) or its ramifications, functioning is not markedly impaired, and behaviour is not obviously bizarre or odd’ (APA 2013 p. 90). In other words, when a subject suffers with monothematic delusion, i.e. a monothematic delusional disorder or syndrome, the specific anomalous cognitive content which the subject thinks about, recounts and otherwise acts upon typically occurs within a ‘culture medium’ of accurate/understandable/predictable enough beliefs and behaviours. Which is to say, the subject generally operates within a remit of normal.

Plus if the delusion with monothematic contents occurs within the context of an existing and perhaps complex psychiatric problem, the basic story stays the same but some particulars may need to be considered with sensitivity. This is because what counts
as the remit of ‘normal’ can and often does change for subjects who have to contend with the complicated circumstances associated with chronic psychiatric illness.

This is a substantive claim. So one might ask is there evidence to support it? And in the process of selecting that evidence we are faced with an elementary choice. On the one hand, we may edge ‘inwards’ in an attempt to isolate and analyse the normality, i.e. efficacy (or otherwise), of, say, the Bayesian reasoning processes which play-out across a subject’s internal cognitive architectures. On the other hand, we can move ‘outwards’ and ascribe qualitative assessments of rationality (or otherwise) to the subject’s observable, i.e. public or social, behaviours. However, to be clear, this inner/outer distinction is a fiction or false dichotomy across a number of vectors – not least of which is that the tests by which subjective reasoning functions are investigated necessarily entail social acts such as instruction-giving, object-watching, button-pushing and verbal declarations in order for a subject to engage with the relevant procedures. Nevertheless, if we choose to draw an arbitrary boundary somewhere in the region of a subject’s skin, then we may use the resulting inner/outer distinction as a handy theoretical tool.

So, concerning evidence relating to the efficacy (or otherwise) of the reasoning processes of subjects who suffer from monothematic delusion, Sullivan-Bissett and Noordhof (2015) identify several studies which indicate that said reasoning processes fall within the remit of normal, and which, thereby, challenge the idea that the explanation of said delusion need/ought appeal to the fact of significant abnormality concerning the machineries of the cognitive system (p. 5). For example, Sullivan-Bissett and Noordhof point to research studies which demonstrate that delusional subjects tend to arrive at Bayesian inferences faster than non-delusional subjects, but that despite the speed with which relative inferences are arrived at, neither group demonstrates a greater/lesser tendency to hold on to a particular conclusion come what may (ibid). If this is the case, then what might be in play is a relative lack of caution on the part of delusional subjects and/or relative over-caution on the part of the non-delusional subjects, but, in any event, the evidence does not point to the presence of deep irrationality in either group.

Likewise, evidence relating to the normative aptness (or otherwise) of public behaviours performed by subjects suffering from monothematic delusions tells against the idea of requisite deep irrationality on the part of subjects who have this sort of problem. Consider, for example, the case of JK, a subject diagnosed as suffering from
Cotard delusion and who, thereby, believed that she was dead (Young and Leafhead 1996). When JK was asked to consider the implication of her heartbeat, her response was that, given the fact that she were dead, her heartbeat was not a sign of life; though she acknowledged that this situation was one that most people would find hard to believe (p. 158). JK's reasoning is valid enough – if she were dead, then, counter to the norm, her beating heart would not be a sign of life. So we can say that JK's Bayesian reasoning processes are functioning efficaciously enough. Yet more than this, as Sullivan-Bissett and Noordhof point out, JK's public or social performance is normal too – at least in so much as JK understands the norm; understands the normal social attitudes and behaviours in relation to the norm; however, given the special/difficult circumstances in which she finds herself, JK reasons that, given the facts of her case, the norm does not apply to her now (Sullivan-Bissett and Noordhof 2015 p. 6). From our perspective, i.e. an objective third-person perspective, yes, there is certainly something amiss. However, what is not amiss is JK's reasoning processes or behavioural performance. Indeed, it would be hard to argue against her logic. Moreover, JK’s resistance to objective third-person reasoning is same in kind to the innumerable resistances expressed by countless other ‘normal’ subjects who find themselves negotiating their own special/difficult circumstances; for instance, consider the wishful-thinking and subjective self-deceptions conjured up in the face of blatant third-person facts to the contrary with regard to feckless, faithless and abusive lovers (Noordhof 2003).

A similar story can be told regarding the case of Mr P, that is to say the subject suffering from perceptual delusional bicephaly whose monothematic delusions occur within the context of a major psychotic disorder (Ames 1984). It is observed that Mr P presents with ‘a history of previous treatment at various psychiatric hospitals over the last six years ... [And it is] significant that his wife had died in a car accident two years previously, and that he was the driver in that accident’ (p. 193). On admission Mr P describes ‘a second head on his shoulder. He believed that the head belonged to his wife’s gynaecologist, and described previously having felt that his wife was having an affair with this gynaecologist, prior to her death’ (ibid). Moreover, Mr P experiences auditory/verbal hallucinations of several different voices, one of which is that of the gynaecologist emanating from the second-head, and all of which (voices) confirm the presence of the head, while the head itself gloats that it is going to take the wife away (ibid).
Ames does not say what the previous psychiatric treatments involved, nor does he say anything about the patient’s circumstances immediately prior to the gunshot incident. Nevertheless, here is a subject who, one would assume, has gone through and remains tangled in some noteworthy special/difficult circumstances. Including the circumstance of a six-year history of (unspecified) psychiatric illness and treatment, and that on admission he is diagnosed as suffering from a major psychotic disorder, schizophrenia. In other words, beyond the problem of the monothematic delusion itself, Mr P’s levels of functioning in one or more key areas of everyday life (e.g. his linguistic or non-linguistic thoughts/behaviours in re interpersonal relations) are clearly derailed. And while the belief about his wife’s infidelity may or may not be a (non-bizarre) persecutory-type delusion or instance of morbid jealousy, it, nevertheless, forms part of his belief-set relating to the circumstances which confront him now. And, again, from an objective third-person perspective, one would not wish to say that sustaining a serious head injury on the grounds of shooting a phantom second head amounts to typical behaviour. Nevertheless, given the special/difficult circumstances with which Mr P is contending, his reasoning/behaviours can be construed as functionally normal, i.e. cohesive and comprehensible (which is not the same as viable, permissible or condonable), in that a line of reasoning (albeit subjective and self-absorbed reasoning) can be tracked; and which, moreover, has much in common with other non-psychotic and psychotic cases of ‘obsessive love’ (a notion which includes, amongst other things, cases of erotomanic delusion, erotomania and stalking behaviours) and which features a long and well-documented history of cuckolded men (mostly men) perpetrating often tragically predictable acts of violence against the significant other and/or the (actual or supposed) paramour (e.g. Meloy 1998; Boon and Sheridan 2002).

The idea that special/difficult cultural, social and interpersonal circumstances are linked to the onset and development of psychosis and psychotic disorder is supported by the findings of numerous studies. For instance, cultural, social and, especially, interpersonal circumstances assume a significant role in occurrences of erotomanic delusion, erotomania and obsessional following (Mayers 1998). Similarly, the delusions of possession syndrome are ‘better understood by reference to socio-cultural background’ (Yap 1960). DSM-5 specifies that as part of the diagnostic process a subject’s cultural background needs to be taken into account and that ‘the content of delusions... varies across cultural contexts’ (APA 2013 p. 93). Furthermore, Richard Bentall addresses the contributory role played by the ongoing occurrence of stressful
socio-cognitive circumstances on psychosis and relapse (Bentall 2003 [2004] pp. 428-34). And the link between early-life adversities and the symptoms of adult psychosis has been revealed to be significant (Bentall et al. 2012; Sitko et al. 2014).

Still in itself, evidence of a link between special/difficult cultural, social and interpersonal circumstances and the symptoms of psychosis does not refute the idea that clinically abnormal cognitive processes have a necessary causal role in cases of delusion. For instance, the empiricist two-factor approach does not preclude the idea that in individual cases of monothematic delusion additional (contingent) factors could be in play. And indeed, an example of such a contingent factor might well be the special/difficult cultural, social or interpersonal circumstances in which any particular subject may or may not be currently involved. Nevertheless, what the empiricist two-factor account does demand is that even in cases where special/difficult circumstances assume what could be viewed as an explanatory role, the subject’s delusion must first-and-foremost be explained in terms of a clinically significant pathology of the neural machineries which realise conscious experiences with delusion-specific anomalous actual contents plus a clinically significant pathology affecting the cognitive functions associated with belief-formation processes re said anomalous experience and which result in the onset and/or development of the delusional belief.

However, it is the case that in very many cases of monothematic delusion, the occurrence of anomalous circumstances has ‘so strong a prima facie connection’ to the incidence and, indeed, the content of the delusion, that they (the anomalous circumstances) certainly seem to assume the role of a significant causal factor (Coltheart et al. 2011 p. 285). Furthermore, as argued above, in some cases at least, once the anomalous circumstances are taken into account, the subject’s reasoning processes and behaviours can be said to fall within a remit of normal, in that they (the reasoning processes) demonstrate a valid line of reckoning and extrapolate to predictable public behaviours, at the very least, from the perspective of third-person observers who are in the know. Moreover, and again as argued above, these public behaviours do not require an explanation in terms of a distinct aetiology, but are ‘simply’ another instance of readily observable commonplace behaviours in which people comprehend and conceptualise the facts re their environmental, cultural, social and interpersonal surroundings and in light of their own subjective needs, wants, desires, experiences and biography, then behave on said comprehensions and conceptualisations accordingly,
even if in some cases, such as those involving the feckless, faithless and abusive lovers, the resultant behaviours are in blatant contravention of third-person facts.

Yet if it really is the case that the public behaviours in question are commonplace behaviours which are amplified by the effects of special/difficult environmental and motivational concerns, then this is a blow to the empiricist two-factor account; since the idea of clinical neuropathology effecting deep irrationality loses much of its explanatory import and may even be regarded as a contingent, as opposed to a necessary, explanatory feature in cases of monothematic delusion – which actually is a likelihood considered then rejected by Coltheart et al. (Coltheart et al. 2011 p. 285).

Indeed, one would not deny that in some cases of monothematic delusion a significant pathology of the cognitive machineries may assume a significant explanatory role. For example, as was previously mentioned, Farnham et al. describe a case of a subject with a worsening history of grand mal epilepsy, impaired verbal memory, and symptoms suggestive of temporal lobe epilepsy; who was arrested, charged and sentenced on account of stalking behaviours; subsequently held under the Mental Health Act, and diagnosed with erotomania; and, then, found to have a large arteriovenous malformation in the left frontal lobe which impinged on the left temporal lobe (Farnham et al. 1997). The subject’s epileptic symptoms were treated by medication, the psychotic symptoms resolved, the lesion was later removed, and twelve-months afterwards no further episodes of stalking behaviour were reported (p. 710). Whereupon Farnham et al. conclude that the ‘case is, apparently, the first report of erotomania and stalking behaviour secondary to an undiagnosed vascular lesion’ (ibid).

Nevertheless, one may query the precise role played by the lesion with regard to this subject’s delusion and psychotic disorder. For instance, this was a large haematoma affecting a sizable area of brain which does not, on the face of it, assume the characteristic profile of the kind of discrete pathology affecting the brain’s cognitive machineries posited by either the empiricist two-factor thesis or (relationist) rationalist thesis. In addition, the subject’s psychotic symptoms diminished on admission at the onset of medical treatment and prior to the removal of the haematoma. And, finally, one cannot simply factor out the effect on the subject of some markedly special/difficult circumstances re arrest, sentencing, prison, detainment under the Mental Health Act, the discovery of a significant medical problem affecting his brain, perhaps seeing the dramatic MRI images, and undergoing surgical procedures to correct the damage.
And even if for the sake of argument we choose to allow that this particular case is a paradigm example of a significant pathology negatively impacting the relevant workings of the brain’s cognitive machineries and so directly effecting the anomalous cognitive contents characteristic of delusion, it still leaves open the problem that in very many other cases of delusion, not only is brain damage of this sort simply not evidenced, but once any special/difficult circumstances faced by the subject are factored in, then explanatory appeals to deep irrationality on the part of the subject become superfluous and/or quite simply fail.

**Appeal to Abnormal Experience**

The kind of abnormal experience in question is essentially a perceptual-type experience which relates to no, or no clearly defined, stimulus in the observable world. The phrase is used to refer to illusions, where a stimulus is there but it is somehow misperceived, and hallucinations, where no significant stimulus is present.

Regarding the role of anomalous experience in delusion, it is recognised that different types of delusion characteristically feature anomalous experiences of different intensities or strengths, which (strengths) in chapter one were glossed as being zero, weak and (very) strong. Thus erotomanic delusion is identified as a case which characteristically involves zero significant experiential element. Capgras delusion as a case which tends to be associated with weak anomalous experiences, such as atypical feelings of affect. And the case of perceptual delusional bicephaly involving Mr P is offered as an example of a delusion involving (very) strong hallucinations of voices and a second head.

Also explained previously is that the empiricist theses on delusion in all cases award a necessary role to the factor of anomalous experience, whereas Campbell’s (relationist) rationalist thesis necessarily rejects the possibility of anomalous experience, qua hallucinations, but nevertheless allows that weaker experiences, such as those of anomalous affect, might have a downstream contingent contributory role.

However, in this section, although the notions of zero and weaker anomalous experiences both feature significantly in the discussion, it is the notion of hallucination in delusion which is of most interest. And the reason for reflecting more closely on hallucination is that experience, and specifically the nature and role of experience in perception and hallucination, is a core theoretical component for each of the prevailing
theories of delusion, plus hallucination, and specifically (very) strong hallucination, is explicitly recognised as a key clinical feature of psychotic disorder. Specifically, from DSM-5:

Hallucinations are perception-like experiences that occur without an external stimulus. They are vivid and clear, with the full force and impact of normal perceptions, and not under voluntary control. They may occur in any sense modality, but auditory hallucinations are the most common in schizophrenia and related disorders. Auditory hallucinations are usually experienced as voices, whether familiar or unfamiliar, that are perceived as distinct from the individuals own thoughts. The hallucinations must occur in the context of a clear sensorium; those that occur while falling asleep (hypnagogic) or waking up (hypnopompic) are considered to be within the range of normal experiences. Hallucinations may be a normal part of religious experience in certain cultural contexts. (APA 2013 pp. 87-8.)

So thinking about hallucination. One way of categorising hallucinations in philosophy is to say that the hallucination itself is weak or (very) strong. A weak hallucination comprises a perceptually short-lived and relatively indeterminate experience. As Dennett describes it, instances of weak hallucination include ‘flashes, geometric distortions, auras, afterimages, fleeting phantom limb experiences and other anomalous sensations’ (Dennett 1991 [1993] p. 7). Alternatively a strong hallucination comprises a long-lasting and definite experience. Or, in Dennett’s words, ‘an apparently concrete and persisting three-dimensional object in the real world’ (ibid). One can also discriminate between inert and dynamic strong hallucinations. For instance, the following hallucinatory dagger Campbell describes is inert since, as realistic as it is, the hallucination just sits there, (seemingly) suspended in mid-air:

Suppose that a [real, actual] dagger is hanging in the air before you, and you are looking at it closely. You are visually attending to it. There is a [real, actual] dagger to which you are consciously attending. What can we say to compare and contrast this with the case in which you are having a vivid hallucination of a dagger, and this hallucination is occupying your attention? Just to be fully explicit, the case I have in mind is one in which the ordinary case of attention to
a [real, actual] dagger is matched as closely as possible by the hallucinatory experience. That is, if the ordinary dagger seems heavy and substantial, so too does the hallucinatory dagger; the hallucination does not, for example, shimmer unduly, or seem any more bloodstained than daggers usually do. (Campbell 2002 pp. 116-7.)

This case of an inert strong hallucination can be contrasted with Dennett’s example of a dynamic and interactive (very) strong hallucination of a ghost:

A strong hallucination would be, say, a ghost that talked back, that permitted you to touch it, that resisted with a sense of solidity, that cast a shadow, that was visible from any angle so that you may walk around it and see what its back looked like. (Dennett 1991 [1993] p. 7.)

As Dennett further observes, reports of strong hallucinations are rare and should always be treated with suspicion, but be that as it may there are numerous reports of strong multimodal hallucinations in the literature of clinical psychology (ibid).

And certainly Dennett’s observation re (very) strong hallucinations in clinical settings is wholly in line with the specification of hallucination in DSM-5. Nevertheless, it needs to be emphasised that not all cases of delusion in clinical settings feature abnormal experiences. For example, as has already been observed and was outlined in chapter one, erotomanic delusion is a case that characteristically involves zero abnormal experience, and should hallucinations feature they tend to be a later development which happens as the overall problem gains in complexity, i.e. is changing to become a more complex psychiatric disorder, but even at this stage hallucinations tend to be dreamy and somatic in kind (e.g. Kennedy et al. 2002). Then again re other cases of delusion (very) strong abnormal experiences clearly feature. These might be hallucinations in a single modality, such as those of hallucinated voices. For instance, once when I was working on the hospital wards, a colleague told how moments before she had entered the dayroom and heard its only occupant, an elderly lady who I’ll call Ethel, whisper (ostensibly to no one, or at least to no one that Val could see), ‘I can’t talk anymore, nurse has come in’. Or they might be rich multimodal hallucinations, such as those of a patient I’ll call Andrew who swore that he could see through time and space. Or they
might be similar in kind to those which characterise the case of Mr P and his very strong hallucinations of the goading belligerent second head.

However, a consequence of this is that, given the mere details of the data set, any blanket proposal that abnormal experience assumes (or does not assume) a necessary explanatory role in any given case of delusion is at best an oversimplification. Yet, more than this, if one pokes the issue just a little bit further, then it becomes apparent that the facts surrounding abnormal experience in all kinds of clinical cases generate all sorts of problems for each of the prevailing accounts.

Take the empiricist theses. On the face of it, the empiricist oughtn’t have a problem accounting for cases of delusion in which anomalous experience of a significant sort is held to occur. However, if we consider the commonkind nature of experience at work in the empiricist model, in relation to the nature of experience as it is held to occur in delusion, then it seems that matters are not quite as straightforward for the empiricist as they first appear.

Think back to the empiricist theses as they were mapped at the start of this chapter. P1 is the premise that experience in perception and hallucination is actual and common in kind. It is the empiricist’s foundational proposal, the starting-point of both empiricist arguments, and as such it can’t be challenged in and of itself. However, as was noted, what can be challenged are downstream consequences of sanctioning P1.

So taking the empiricist one-factor account. P1 of the empiricist one-factor argument is that consciousness (experience, awareness) rests on its own set of neural correlates, that actual experience occurs at the end of a causal chain, and that said experience is commonkind in veridical perception and hallucination, with the difference being that in hallucination the causal chain sparks independently of a significant stimulus in the observable environment. If we add to this P2, which is the largely accepted view that actual commonkind experience is effectively made sense of by the cognitive system, the workings of which rest on a discrete set of neural correlates, and via which cognitive workings one arrives at one’s beliefs about the world. Next is the observation that, by and large, normal experiences, give rise to true beliefs (P3), or sometimes to understandable false beliefs (P4). And delusion is a fundamentally normal though by definition false belief (P5). And so with no warrant to posit a deep abnormality of cognitive functioning (P6-P8), for the one-factor theorist, an abnormal experience must be where the key explanatory factor lay (C).
In addition, it is fair to think that what is at point at C is not just any abnormal experience, but a markedly or significantly abnormal experience. The reason for this assumption being that if the abnormal experience in question were merely a trifling or even a middling irregularity, then while one might agree that a discretely functioning cognitive system which is *ex hypothesi* normal might, in such circumstances, deliver, say, a short intellectual jolt or bit of passing perplexity or brief episode of dissociation or fleeting anomalous salience in response to said experience, on the whole, such a cognitive system could hardly be so temperamental as to be thrown into such serious disorder as results in a diagnosis of severe psychiatric illness for the subject – not least because, if this were the case, then the symptoms of psychiatric illness would actually be the (or, at least, a viable) norm.

Thus if the empiricist account is a one-factor approach, then reasonably it can be summed up as claiming that a significantly abnormal experience is realised deep in the neurological underpinnings of consciousness, and is subsequently judged by a discrete and normally functioning cognitive system via a logical development which, so to account for the anomalous experience, eventually settles on a markedly odd conviction which (potentially) warrants a clinical diagnosis of delusion or delusional disorder. However, if it is looked at from the perspective of clinical delusions, this proposal runs into difficulties.

First, if we take the causal commonkind theory of perception and hallucination. In veridical perception, experience is said to be the actual (definite, real) ongoing (progressive) unified (polished) emergent property in (the workings of) the brain, which (experience) marks the termination of a train of causal changes which starts with the fact of an discernible stimulus in the subject’s observable world. Conversely, in hallucination, while the outcome in experience is identical in kind to that of veridical perception, deep in the workings of the brain, the causal chain which realises said outcome sparks up independently of a significant discernible stimulus. Hence the experience for the subject is that of a hallucination as opposed to that of a perception.

Yet, Dennett observes, when a brain shorts its own circuits of experience (or gets addled its machineries involved in experience), the result with regard to subjective experiences tends to be only the weakest and most fleeting of hallucinations (Dennett 1991 [1993] p. 9). For instance, the simple visual hallucinations of elementary objects and/or repeating patterns which are characteristic of occipital lobe pathology (stroke, seizures) and migraine (Knight *et al.* 2008); or perhaps episodes of incongruous affect
(motivation) and feeling (emotion, mood) (Ellis and Young 1990); or the sorts of brief somatic hallucinations, such as tickles and shooting pains which are characteristic of phantom limb syndrome (Dennett 1991 [1993] p. 8). However, it is hard to see how these sorts of simple or weak anomalous experiences could function as the markedly or significantly anomalous experiences which are held to be the sole necessary and sufficient explanatory cause of delusion or delusional disorder, simply because a discrete and normally functioning cognitive system would efficiently constrain the glitch.

Still, perhaps it might be argued that Dennett is being overly harsh in claiming that only the weakest and most fleeting of hallucinations would be triggered, since it is known that changes (typically, damage or pathology) in various bits of the brain or perceptual organs can trigger strong anomalous experiences in various modalities. For example, the hallucinations of scenes, figures, animals etc which can occur in cases of midbrain pathology, Parkinson’s syndrome or certain kinds of dementia (Knight et al. 2008); or the vivid visual hallucinations, with varyingly mundane, odd or wondrous (though characteristically non-threatening) content relating to unfamiliar figures (often children, and children at play), creatures, objects and scenes, which are indicative of Charles Bonnet syndrome (ffytche 2013; Knight et al. 2008). It is fair to say that a vivid visual hallucination of lively children would be classifiable as a significant anomalous experience. Nevertheless, appeal to these sorts of stronger experiential anomalies as support for an account of delusion simply generates further issues and problems for the (one-factor) empiricist.

One problem for the empiricist is that the incidence of such experiences cannot be separated from (numerous) wider factors, such as sleep disorders; the sudden and severe loss of visual acuity; neurological problems, such as occur in Parkinson’s and dementia, plus the many health, personal and social problems which can arise as a result; and wider social factors, such as the subject’s background circumstances, understandings, expectations etc (Knight et al. 2008; Kapur 2003; Bentall et al. 2012; Sitko et al. 2014). This is certainly not to suggest that the reported strong anomalous experiences do not occur, since plainly something is going on and, moreover, it is something which can have substantive personal and clinical significance. However, not only is it by no means certain that the fundamental nature of said happenings is that of a neurologically discrete, actual, ongoing, polished phenomenal experience which occurs wholly (or even largely) upstream to cognitive processing, but the evidence actually tells against this idea, since that which is accessed, reported, recorded and assessed by the
patient, family, clinicians etc is plainly an interweaving of very many different biological, affective, emotional, intellectual, personal, social and cultural factors (ibid.). That is to say, it is essentially impossible to definitively tease the fact of the matter apart. Consider, for instance, the quotation above from DSM-5 which observes that hallucinations may be a normal part of religious experience in certain cultural contexts (APA 2013 p. 87-8). And, certainly, the role of wider factors as regards patient reports of hallucination is a state of affairs with which health care professionals regularly contend. For example:

When patients exhibit VH [visual hallucinations] and other neuropsychiatric phenomena, appropriate sensitivity to the role of cultural factors is an important determinant of the success of the medical consultation. (Knight et al. 2008 p. 103.)

Hallucinations are a salient example of an entire gamut of neuropsychiatric phenomena in which social and cultural factors can powerfully determine the success or failure of the medical consultation, and these factors will become increasingly relevant as our society becomes globalised. (Knight et al. 2008 p. 105.)

Thus to hold the rigid view that a strong abnormal experience (of, say, a visual hallucination of children playing or a pastoral landscape or a suspended dagger) is (by necessity) the sole significant explanatory factor in any given case of delusion or delusional syndrome is to determinedly select out many other factors, and any or all of which could be equally if not more explanatorily significant to a particular case.

A second and not unrelated problem for the empiricist is that the sort of strong anomalous experience in question is not routinely associated with cases of psychosis and psychotic disorder. Think back to the above extract from DSM-5. Not only is contextual sensitivity required concerning hallucinations which are held to occur as a normal part of certain religious practices, but hallucinations which occur around sleep are by definition not a recognised feature of psychosis or psychotic disorder. Recounts of sleep-related hallucinations tell of illusions, visions, phantasms etc occurring in various perceptual modalities and ranging from weak dreamy occurrences to vivid threatening strong occurrences (e.g. Sacks 2012 [2013] pp. 199-228). That is to say, they constitute a range of experiences which, from the point of view of the subject and
in some cases at least, are classifiable as a marked or significant anomaly. Nevertheless, from the psychiatric perspective, in and of themselves, sleep-related hallucinations are explicitly held to fall ‘within the range of normal experience’ (APA 2013 p. 88).

Likewise, one may assume that the same applies to the strong hallucinations which occur in conjunction with certain wider circumstances, such as, say, the sudden and severe loss of visual acuity which is characteristic of Charles Bonnet syndrome. According to the Royal National Institute of Blind People (RNIB):

CBS [Charles Bonnet syndrome] hallucinations are only caused by sight loss and aren’t a sign that you have a mental health problem ... Your hallucinations can be of almost anything you can think of – they can range from simple patterns, shapes or colours, to vivid detailed pictures of people, animals, objects or buildings ... Usually with CBS you’re aware – or can learn to recognise – that what you’re seeing isn’t real even though it’s very vivid. People with CBS don’t usually develop complicated thoughts or ideas about why they are seeing things. (Online. https://www.rnib.org.uk/eye-health/eye-conditions/charles-bonnet-syndrome-cbs. Accessed 17 September 2019.)

Charles Bonnet syndrome hallucinations only ever occur in the visual modality. And while in some cases the content of the hallucination may be so mundane and lifelike that the subject may not (initially) realise that he or she is hallucinating, hallucinatory contents can also be fantastic or frightening. For instance, consider having a (repeated) seemingly veridical experience of an unknown person or persons of realistic appearance in one’s home (ibid). Hence, from the point of view of the subject, and in the early stages of the disorder at least, Charles Bonnet syndrome hallucinations can unquestionably involve marked and significant abnormal experiences. Nevertheless, Charles Bonnet syndrome is not a psychiatric disorder. Subjects who experience the hallucinations typically do not develop complex or delusional ideas in order to explain the occurrence and/or content of the hallucinations. And neither is there any suggestion in the literature that subjects who undergo Charles Bonnet syndrome hallucinations have any sort of tendency to have or to develop associated psychiatric problems.

All of this is problematic for empiricist accounts of delusion, simpliciter. Yet specifically still thinking about the empiricist one-factor approach, its defining claim is that clinically abnormal experience is the single significant necessary and sufficient
explanatory cause of delusion. Yet here is a range of cases in which anomalous experiences of a marked and significant nature are recognised to occur, but which anomalous experiences are of themselves clearly insufficient to effect delusions or a delusional disorder. Indeed, the antithesis of the one-factor thesis seems to be occurring. For if we accept the empiricist explanatory model re the relationship of experience (P1) to cognitive processing, i.e. the forming of factual beliefs (P2), then given the significant occurrence of strong hallucination, in conjunction with the significant non-occurrence of psychosis or psychotic disorder, so the evidence not only supports the view that abnormal experience cannot be the sole significant (i.e. necessary and sufficient) explanatory cause of delusion and delusional syndrome, but it supports the observation made above that a normally functioning cognitive system is not markedly temperamental, but is typically more effective at constraining experiential oddities than the empiricist one-factor account of delusion can allow.

A third problem for the empiricist is that although the kinds of experiential anomalies which occur in cases involving midbrain pathology, or Charles Bonnet syndrome, or which happen around sleep are often reported to have a strong complex content, characteristically all such anomalous experience is (effectively) inert. That is to say, a characteristic of such cases is the absence of interaction between the subject and the content of the hallucination and/or between the content of the hallucination and the surrounding world. For example, the hallucination is: motionless, it just stands or hangs there suspended; has limited movement, for instance, the subject hallucinates a large spider ascending the wall; maybe has more by way of movement, for instance, a hallucination of lively albeit silent children; perhaps involves a tactile feeling of pressure or visceral threat, for instance, ‘I wake on my stomach to the sensation that the mattress is breathing. I cannot move and the terror sets in …’ (Sacks 2012 [2013] p. 220); or possibly has some basic linguistic content, for instance, ‘all of a sudden he [the hallucination] started shouting out random numbers, like FIVE-ELEVEN-EIGHT-ONE-THREE-TWO-FOUR-ONE-NINE-TWENTY, then laughed hysterically …’ (p. 224). Essentially, these are actual examples of hallucinations which are comparable to Campbell’s theoretical example of a solid-looking seemingly blood-stained dagger suspended in mid-air. And certainly the (effectively) inert nature of some of these hallucinations does not detract one bit from their being marked or significant occurrences. For instance, a suspended dagger would be momentous enough, but if in the supposed security of one’s home or one’s bed one turns to see an unanticipated
stranger (be they sickly green in hue, in full native dress, bearing angel wings or devil horns, or entirely commonplace in appearance) standing or laying beside one, such an experience would be classifiable as marked or significant in anyone’s book (Sacks 2012 [2013]). Nevertheless, abnormal experiences of this sort still fall short of the complex dynamic continuing hallucinations evidenced in many cases of psychosis and psychotic disorder, such as those featuring companionable hallucinatory voices or a belligerent and goading second head with which the subject explicitly interacts, and which are closer in kind to Dennett’s theoretical example of a very strong multimodal hallucination of a three-dimensional shadow-casting ghost.

However, now the problem for the empiricist is that while goings-on involving very strong, complex, dynamic and continuing abnormal experiences are associated with cases of psychosis and psychotic disorder, given the empiricist explanatory framework re the causal commonkind theory of perception and hallucination, it is hard to conceive precisely what change to the causal machineries of a discrete perceptual system could possibly effect the sorts of ongoing developing dynamic interactive multimodal anomalous occurrences which are a central feature of at least some delusions and delusional disorders. Indeed, it seems implausible to think it could do so.

Nevertheless, an empiricist response to this objection might be to say that we know that neurological damage can effect perception such that it realises simpler kinds of hallucination; and in some cases we also know that such damage is associated with stronger albeit inert hallucinations; so while we currently do not know precisely where lays the organic malfunction which effects the kind of strong, complex, dynamic etc hallucinations which are characteristic of some psychiatric disorders, and specifically of some delusions and delusional disorders, we can nevertheless reason inductively, on the basis of what we do know, that such damage is ‘in there’ somewhere, just waiting to be pinpointed, as it were. Thus far from being implausible the proposal that a relevant change to the causal train of a discrete perceptual system effects the sort of ongoing developing dynamic interactive multimodal anomalous hallucinations which are characteristic of some delusions and delusional disorders is clearly possible in principle.

In turn, one response to this empiricist line of reasoning would be that afforded by Dennett, which is basically that, no, it is not possible in principle (Dennett 1991 [1993] pp. 3-10). Just because we can conjure up a notion in imagination does not for a moment necessitate that the details of that notion are possible in fact. And given the principle of the commonkind thesis, and given the criteria of (very) strong
hallucinations, so (very) strong hallucinations are impossible. This is because if hallucinations are matter of a brain launching an ungrounded causal train in its perceptual workings, then while transient luminosities and sensations might be effected, once anything more complex becomes a factor, and certainly once dynamic interactions with or by hallucinatory contents are said to feature (pp. 4–6), so the notion of commonkind hallucinations becomes ‘simply impossible’ (p. 7), not least because of the sheer informational explosion which would arise tantamount immediately if a brain were to progressively engage with the toing and froing of spontaneous responsive behaviours as said behaviours unfold in the world (p. 9).

[T]here is no way the brain as illusionist could store and manipulate enough false information to fool an inquiring mind. (Dennett 1991 [1993] p. 9.)

Furthermore, neither can the empiricist make judicious appeal to the idea of the cognitive penetration of experience – i.e. the idea that the content or character of perceptual experience can be penetrated or affected by non-experiential contents, such as by a subject’s existing factual beliefs (Siegel 2010, 2012) – to account for delusions or delusional syndromes in which a (very) strong hallucinatory element seems to assume a central explanatory role. This is not to offer comment on whether the cognitive penetration of experience does or does not happen. It is simply an observation that, if the cognitive penetration of experience were held to be the case, then while the fact of cognitive penetration may well go some way to explaining the incongruous contents of some strong hallucinations as well as making available some sort of selective or anticipatory function regarding the informational contents necessitated by responsive behaviours, perhaps the most obvious of which would be the ostensible dynamic interchanges of linguistic content, it also generates a deep and destabilising uncertainty as regards the empiricist/commonkind model. This is because, if experience were so penetrated as to allow for the occasioning of dynamic interactive perception-like hallucinations of things like true-to-life daggers, visitations by angels, incorporeal interlocutors, and a belligerent second head, then it seems reasonable to think that the role of cognition in experience per se could be more deeply substantial than the causal commonkind theory of experience can feasibly allow, for the reason that the prospect of such occurrences generates uncertainty concerning the point at which experience stops being the upshot of specific causal changes realised in the discrete neural correlates of
consciousness and becomes instead a predominantly or even fundamentally cognitive affair, with the ramification being that if experience is said to be largely or even wholly cognitive in nature in the case of (very) strong perception-like hallucinations, then might it not possibly be the case that experience is largely or even wholly cognitive in nature in the case of humdrum perception too?

Yet in addition to the problems generated for the (one-factor) empiricist concerning delusions in which strong anomalous experience, and especially very strong anomalous experience, certainly seems to assume a central explanatory role, there is also the reverse difficulty for any empiricist of explaining those cases of delusion which certainly seem to lack a significant explanatory element, for instance erotomaniac delusions. And while this may seem an uninteresting weakness to point out, it is actually a non-trivial concern. Not least for the reason that, if delusions involving (very) strong hallucinations, such as that of a belligerent second head, are relatively rare, the opposite is true of delusions which typically have zero significant experiential element, and which class includes delusions of reference, paranoid or persecutory delusions, and delusions of grandeur, which along with erotomaniac delusions, are all delusions with which patients regularly present (e.g. APA 2013 pp. 87, 90-3).

What is more, this is not merely a case of drawing attention to a counterexample from the clinical dataset, i.e. a class of delusions which the empiricist approach to delusion on the face of things cannot address. (Though this should be borne in mind.) The more interesting concern is that the substantive fact of non-experiential delusions draws attention to deeper theoretical concerns which develop should the empiricist maintain a firm hold of P1 – namely, the causal model of experience wherein experience is actual, ongoing and unified, and experience and hallucination are common in kind. The idea is that if P1 is left intact, then a knot of problems develops re P2 and P5 such that C is undermined.

So, leaving P1 intact, attention naturally enough shifts to P2. P2 expresses the idea that (factual) belief follows (actual) experience. It is fair to say that this is the prevailing school of thought re belief. However, as is evidenced by discussion of Dennettian theory in chapter one, it is but one construal of belief, which is to say, it is not unassailable fact. Thus we are quite at liberty to query: Really? Always? And in response to this query, and in sketchbook theory at least, we would basically receive one of two replies. The replies being: yes or no.
So continuing with the sketchbook approach – by which I mean a rudimentary two-dimensional analysis the purpose of which to demonstrate a problem. If the answer one receives is that, yes, this is always the case, then the integrity of P2, and with it the integrity of the empiricist model, is maintained and attention moves on to P3 etc. However, if the answer one receives is that, no, this is not always the case, then not only does P2 fail, but we are left wondering how and why it fails. For instance, perhaps there is a fundamental disjunction at the level of belief, such that some beliefs standardly follow actual phenomenal experiences, for example, Jill looks out of the window, sees the conditions outside, and so formulates the factual belief that it is raining, while some other beliefs do not exhibit this sort of internal logical consistency, but instead might be instances of, for example, Jill’s brain processing multiple streams of basic information and therein effecting a more heuristic functional/behavioural output involving her picking up an umbrella and, if probed, some linguistic output including the word ‘rain’. So, okay, it’s a possibility. However, we would now want to ask in what circumstances might this disjunction at the level of belief occur. Might it perhaps happen in cases of delusion? In which case, our attention moves on past P3 and P4 to P5.

P5 of the empiricist argument is the premise that delusional belief is like or common in kind to any other belief. That is to say, there is nothing fundamental setting a delusional belief apart from the true beliefs of P3 and false but understandable beliefs of P4. We have, however, arrived at the premise via two routes. The two routes being: (i) (factual) belief always follows (actual) experience; and (ii) there is a fundamental disjunction at the level of belief. Either way, we want to know if P5 holds. However, the empiricist may want to distance themselves from (ii). Perhaps they did not like the idea of a disjunction at the level of belief to begin with since it could possibly undermine the integrity of P2 and therein threaten the reliability of P3, P4 and with it P5. Thus the empiricist may prefer to discard that particular line of reasoning. At which, the response to the question would be that, no, (ii) does not apply in cases of delusion. Consequently we are left with (i) delusional belief is common in kind to all other belief, where belief is of the standard construal, i.e. a propositional state which follows actual experience. And so the integrity of P5 is (seemingly) maintained, and with it the empiricist account of delusion thus far, and with it attention moves on to the rest of the argument.

If the remaining argument is that of the empiricist one factor-theorist, then the following three premises (P6–P8) concern the factors of pathology and rationality, which is to say they cover ground already discussed. Accordingly, we arrive at C, which is the
conclusion that given the prior assumption that there needs to be a critical abnormality deep in the biological machineries of the brain, and given the idea that (in monothematic delusion at least) there is no warrant posit deep irrationality, so the culpable factor must be experience, which is to say, in cases of delusion an abnormal experience must be the sole significant explanatory cause.

However, at this point we might want to exclaim, Whoa! But what about all those cases of delusion or delusional syndrome which certainly seem to have only a weak or even zero experiential element? In the case of a weak experiential element, it has already been observed that given a discrete and normal-enough cognitive system there is no real reason why said system could/would not constrain the abnormality. And in the case of no experiential element, the question is how and why can these sorts of delusion even begin to be accounted for given P2, i.e. the standard construal of belief? This does not add up. Thus something in the empiricist one-factor argument must be amiss. And scanning back up (and setting aside P6-P8) attention relights on P2 and/or P5.

P2 is in effect another starting premise. Although it has already been noted that it is certainly not an inviolable truth as there are other equally valid construals of belief on the table. Also there is the logical point explicated above that the idea that (factual) belief always follows (actual) experience might not (always) reflect what is actually going on in re (any particular) belief. All the same, given its role as a starting premise, for now, we can let P2 lay.

In contrast, P5, the premise that delusional belief is like or common in kind to all other belief, it is not a starting premise, but is arguably a (if not, the) central proposal in the one-factor empiricist argument. And, on last pass, we were lax in so much as we simply took the empiricist at their word as regards its application. But not so this time. This time we adopt a more critical stance and ask: Really? Always?

If once again the empiricist maintains that, yes, it is always common in kind, this time we require more than simple assertion. A workable explanation is required as to how and why this is the case. Yet, on the face of things and in all fairness, nothing definitive springs to mind. Accordingly, if the empiricist were to accede that, okay, perhaps in some cases of delusion belief is not, or is not always, common in kind, that perhaps, say, something is happening on the lines mentioned above, so that, concerning delusions or delusional syndromes in which abnormal experience is a (very) significant factor, things progress in the standard, i.e. commonkind, manner. Whereas concerning
delusions and delusional syndromes with only a weak or zero experiential element, then perhaps, and at some point at least in the development of the delusion or delusional syndrome, the belief in question is a non-doxastic state. In other words, there is a now disjunction. Not this time at the level of belief per se, but a disjunction at the level of delusion and which may be limited to, say, the early stages of the disorder and/or specific kinds of delusion.

On the upside, positing such a disjunction allows for the maintenance of the overall empiricist standpoint while accommodating the difficulty of delusions or delusional syndromes with weak or zero experiential content. However, the advantage is a dubious one, for it is only achieved at the cost of diminishing or even wiping out P5 of the empiricist argument regarding delusion, since delusional belief is no longer (or, no longer necessarily) on a par with the standard construal of belief. Furthermore, from this point on, problems would compound for an empiricist taking this route.

To begin with, those cases of delusion which characteristically feature only a weak or zero experiential element are, as observed above, commonly occurring delusions, which is to say, they form a substantial percentage of data set. Thus the empiricist would now be in a position where they would have to accede that in a substantial percentage of cases delusion or delusional syndrome is or is at root a non-doxastic state, which given the definition of delusion as a fixed false belief, not only undermines P5 but also potentially casts a shadow over P2, since it demands that we get very clear as regards precisely what we are talking about when we use the everyday term ‘belief’ and whether, in light of this clarity, clinical delusion is actually a ‘belief’ (or not) and, indeed, whether the standard construal of belief is actually a ‘belief’ (or not).

In addition, the empiricist has the further problem of accounting for delusions and delusional syndromes in which a (very) strong experiential element certainly seems to assume a central explanatory role. This is because, as discussed previously, while delusions and delusional syndromes in which (very) strong experiential contents are said to be a significant feature comprise a relatively smaller percentage of the data set, in the cases in which they are said to feature, the (very) strong experiential anomalies cannot be separated from wider factors, such as the various special or difficult circumstances with which the subject contends, and in the case of very strong interactive hallucinations, while it is hard to conceive precisely what brute changes re a damaged perceptual system could effect these symptoms, it seems that a cognitively-leaning account could maybe go some way to clarifying things.
So, all things considered, the outcome so far of this sketchbook revision is that things aren’t looking good for the empiricist one-factor account. However, once we re-arrive at C matters only get worse. To recap, as regards the empiricist one-factor account of delusion, C is the conclusion that significantly abnormal experience is the necessary and sufficient explanatory cause of delusion. Yet now the question is simply: Why? In the case of delusion or delusional syndromes with a weak or zero experiential element, the empiricist has already allowed that non-doxastic factors such as functional, behavioural or cognitive features are key. In addition, as regards delusions or delusional syndromes with a (very) strong anomalous experiential element, in which wider factors cannot be discounted and cognitive factors are plausibly doing at least some of the explanatory work, then could it not be the case that functional and/or behavioural and/or cognitive and/or and wider factors assume a significant role in these cases as well or even in all cases? Indeed, where is the justification for awarding a central and necessary role to actual anomalous experience? Why not appeal to functional/behavioural/cognitive/circumstantial factors as well or even alone?

Of course, this is only a sketch, a simplified story. And it is a simplified story which an empiricist would likely find reasons to challenge, vociferously. Nevertheless, it is a plausible-enough account which serves to illustrate that, first of all, despite its apparent straightforwardness, the empiricist one-factor approach actually calls on the theorist to take (or not) various logical pathways, at least some of which could entail a substantial theoretical cost, and so which need to be spelled out and examined more closely at some point. And second, it highlights the fact that the rigid and concrete role which any empiricist demands of abnormal experience, generates substantive concerns in terms of both empirical problems, such as the failure to meet a substantial percentage of the clinical dataset, and logical problems, such as a slide towards function and/or behaviour and/or cognition and/or wider circumstantial factors assuming a role as a/the centrally significant explanatory cause.

Still, at the same time, the story I have told is just that, a story, and, moreover, different stories could be told. For instance, following the initial arrival at C and subsequent tracking backwards up through the argument, one could legitimately pass over P5 and so come down on P2 and/or P1 and at that point propose a re-evaluation of belief and/or experience. Although the cost of this could be the pulling down of the empiricist school of thought rather than, as occurred above, a cagey amendment to it at P5.
For instance, one might propose a milder construal of experience and/or belief than that of either the causal theory of experience and/or the standard construal of belief. And such a revision might shift matters closer or even close to the position on experience and belief held by Dennett. And although many empiricists would surely balk at any such revision of their foundational principles, the different way of thinking about things might plausibly open the way for an account of delusion which has a more comprehensive explanatory reach across more of the data set while avoiding the logical pitfalls re experience which confront the prevailing empiricist accounts.

Nevertheless, the next question has to be is the empiricist two-factor account better placed than the empiricist one-factor account with regard to squaring the assumption that in any case of delusion by necessity anomalous experience is always the primary significant explanatory cause? And, in short, the answer is that, no, not really, it isn’t.

To recap, the empiricist one-factor theorist’s position is that anomalous experience is the sole necessary and sufficient significant explanatory factor in any given case of delusion, whereas the empiricist two-factor position is that upstream abnormal experience in conjunction with downstream cognitive processing are the individually insufficient but jointly necessary explanatory causes in any given case of delusion. On the face of things, the positing of the second explanatory factor might seem to afford the two-factor theorist an advantage since, if the one-factor theorist faces difficulty explaining how, given the constraints of a normally functioning cognitive system, a weak experiential anomaly could achieve explanatory significance in cases of delusion, so by positing a further deep abnormality of the cognitive system, the two-factor theorist weakens said constraints and so ostensibly offers a reason as to why following anomalous experience delusion could result. Nevertheless, on closer examination of the two-factor argument, it becomes apparent that, at best, the role of experience cannot be necessary but is at best contingent in cases of delusion, as well as there being a further drawback regarding a slide towards cognitive factors assuming the central, or maybe in some cases even the sole, explanatory role.

The problem for the two-factor theorist centres on the way in which the two sides of the theory are presented and then required to work together as a single account. In a nutshell, first, the two-factor theory makes foundational appeal to the basic empiricist position, as this is outlined in P1 to P2 above. Second, it observes, plausibly enough, that as things stand the basic empiricist position is insufficient as, on the whole,
a normal cognitive system would work to constrain an experiential anomaly (P4-P7). Thus, from this, the two-factor theorist concludes that, subsequent to a prior and necessary (clinically) anomalous experience the workings of a deeply damaged cognitive system, i.e. a (clinically) abnormal belief formation process, must also assume a necessary explanatory role (P8-P10 and C). And, at this juncture, the two-factor theorist may point to a case of delusion which characteristically features a weak experiential anomaly, such as a transient pain or tingle or feeling of anomalous effect, and the content (character) of which anomalous experience is purportedly seized on by the proposed deeply damaged cognitive system, so that, said system, being deeply damaged, fails to have or apply well-functioning constraints to curb its reckonings, with the result that cognitive anomalies gain traction, with the result that abnormal downstream effects characteristic of the delusion or delusional syndrome start to surface in the subject’s thoughts and public behaviours. Capgras delusion is the case which is often used to illustrate the argument (e.g. Coltheart et al. 2010; Coltheart et al. 2011). And so the two-factor theory appears to secure a substantial gain over one-factor theory in that it is able to provide an explanation for delusions which exhibit a weak experiential factor, whereas arguably the one-factor theory struggles to describe how a weak experiential anomaly could plausibly take hold.

Nevertheless, even if for the sake of argument we allow that the empiricist two-factor account is able to tender a satisfactory explanation for delusions and delusional syndromes which exhibit a weak experiential element, it gains no ground whatsoever over the one-factor account as regards cases which exhibit zero experiential anomaly – unless, that is, the two-factor theorist wishes to say that a deeply malfunctioning cognitive system seizes normal commonplace experience which therein it gets so seriously wrong that a clinically (i.e. psychiatrically) diagnosable abnormality takes hold. In which case, in the case of zero delusions at least, anomalous cognitive processing is now the sole necessary and sufficient factor. In which case, in the case of zero delusions at least, the empiricist thesis is undercut and something which seems to present as a species rationalism re delusion is tended in its place.

And neither does the empiricist two-factor approach gain any ground over the one-factor approach as regards cases which exhibit a (very) strong experiential factor, since one must continue to ask how, within the explanatory remit of the causal theory of perception and hallucination, could strong interactive hallucinations possibly occur – unless, that is, one wishes to appeal to the occurrence of cognitive penetration, perhaps
via a now doubly non-virtuous feedback circle, though, as pointed out above, if actual commonkind experience is penetrable to the extent that cognitive content (normal or otherwise) could effect actual ongoing perception-like visual, auditory and somatic hallucinations of a belligerent second head, then this not only throws into question the extent or veracity of actual experience as an explanatory factor in cases of delusion, but it throws into question the foundational principle of the causal commonkind theory of perception, since, in any instance, cognitive factors (normal or otherwise) could, may, or must be shouldering much, most, or all of the explanatory load.

Alternatively, one might appeal to the idea that the subject mistakenly thinks (i.e., misconstrues) that they see a second head, or that the subject’s discourse with his or her auditory hallucinations is the effect of the same sort of cognitive misconstrual, but in this instance concerning the subject’s own streams of notional/linguistic thoughts running to and fro. And, in fact, this latter notion is not infrequently explored as being the case (e.g. Stephens and Graham 2000 [2003]; Bentall and Varese 2013).

However, on reflection, it might seem somewhat odd that anyone who holds an empiricist view of delusion would elect to say such a thing. For not only does the prospect amplify the role of cognition in delusion, but it also amplifies the role of cognition in experience per se, since if it is possible for one to be so very mistaken in the case of (very) strong hallucination, then, at any given time, we could never be sure via ‘experience’ alone whether we are undergoing the actual experience that we seem to be undergoing, or merely thinking that we are undergoing this actual experience, or actually undergoing any actual experience, simpliciter. And in response, it would do the two-factor theorist little good appealing to the role of deeply anomalous cognitive processing by saying something like: The non-delusional subject has a normally functioning cognitive system and so would never make such a mistake, whereas the delusional subject has a deeply altered/damaged cognitive system which makes such a mistake not only possible but a typical feature of certain psychiatric disorders. For the reason that, while this explanation might mediate concerns re the role of cognition in the non-delusional case, in the case of delusion it still leaves abnormal cognitive processing carrying the brunt of the explanatory load.

So if both empiricist accounts of delusion, when required to address the wider dataset as opposed to specific examples taken from the dataset, are limited in application and encounter a slide towards a cognitive explanation, then, one might ask, does the (relationist) rationalist account of delusion fare better?
However, once again, the answer is no, not really, it doesn’t. Not in this case because there is a slide towards cognition, for, according to the (relationist) rationalist approach to delusion, deeply abnormal cognitive processing is by necessity the sole explanatory factor in play, but, in virtue of the factor of (clinically) anomalous experience, and specifically regarding the role that experience is or is not held to play in specific delusions and delusional syndromes. The conclusion being that while, on the face of it, the (relationist) rationalist approach can account for some of the data set, when required to meet the features of the wider data set, the (relationist) rationalist thesis repeatedly confronts either a substantial need for clarification, or else fails to address what seems to be going on, or else deep sceptical possibilities (again) take hold.

So once more dividing the clinical data set re delusion into three broad categories – namely, those which typically and certainly seem to feature zero anomalous experience; weak anomalous experience; and (very) strong anomalous experience. And taking the class of delusions which characteristically feature zero experiential element first. Then an example of this sort of delusion is erotomanic delusion, which is a disorder which involves a subject maintaining the false conviction that another person (the object) is in communication with and in love with them (the subject), and so is typically glossed as having the delusional content that: S/he loves me; but which does not involve anomalous experience as a characteristic feature, though, in the later stages of the disorder weak dreamlike hallucinations are sometimes reported, alongside the distinctive patterns of behaviour which, should they become public, may well place the subject on the clinical map (e.g. Kennedy et al. 2002).

Thus, in this type of case, the account of delusion afforded by the (relationist) rationalist in terms of (clinically) abnormal cognitive processing – which is to say, in terms of an erroneous framework proposition which stands outside correctness testing, but which operates as a background assumption or norm against which testing takes place, and which thereby potentially affects downstream experiences and/or behaviours, and of which the content s/he loves me could conceivably be a case in point – may seem like a plausible enough construal. For instance, if the subject or the subject’s system repeatedly asks itself the question: Does s/he love me? and repeatedly issues itself the reply: Yes, s/he loves me, then it is reasonable to suppose that under these circumstances such content could eventually influence patterns of downstream behaviour indicative of erotomanic delusion or erotomania and – in line with the
underlying relationist theory of perceptual experience – the dreamy thought-like hallucinations reported in some cases of erotomanic delusion too.

There are, however, problems with the (relationist) rationalist account as it is outlined above, not least of which is that if we take the typical case of an erotomanic subject the object of whose delusion is a prominent celebrity, then it appears to be the case that the flagrantly false content that \( s/be \ loves \ me \) needs not only to be in place, but also to have achieved the status of a deeply fixed framework proposition, prior to the appearance of the fixed false delusion that \( s/be \ loves \ me \) and its associated symptomology. In other words, the relationist account doesn’t so much explain erotomanic delusion as simply shift the bump in the carpet regarding an account of erotomanic delusion, as we now need to know how and why the framework proposition that \( s/be \ loves \ me \) gets going and takes hold.

This may or may not be an intractable problem for a rationalist. Indeed, I discuss from the perspective of a rationalist Dennettian and Prediction Error model how and why in the case of monothematic erotomanic delusion the characteristic contents of the delusion could begin and develop as a result of noise (see chapter four). Nevertheless, the point stands that while, on the face of it, delusions and delusional syndromes which typically feature zero experiential content, but which may develop to include weak or dreamy hallucinations, may seem to be neatly explicable via the (relationist) rationalist account in terms of (clinically) anomalous cognitive processing alone, in the absence of further clarification concerning why and how said anomalous cognitive content gets going, takes hold, and assumes the status of a framework proposition, it appears that matters are not so neatly explainable as they may at first seem.

So, if delusions which feature zero significant experiential content are not clearly explainable within the (rationalist) relationist framework, then what about those which feature a weak experiential element, and of which Capgras delusion is regularly discussed as a case in point. The literature tells us that the weak experiential element associated with Capgras delusion is that of abnormal perceptual saliences and feelings of anomalous affect (e.g. Ellis and Young 1990), while the central criterion of Capgras delusion is that a subject maintains the delusional conviction that a person known to them has been replaced by a simulacrum (imposter, replication) of some sort. The simulacrum may be acknowledged to be visually and behaviourally identical to the
original spouse, child, colleague etc. Nevertheless, the subject maintains that they just know that the person before them now is a fake of some sort (e.g. Alexander et al. 1979).

In a nutshell, Campbell’s (relationist) rationalist explanation for Capgras delusion is that, in any given case, working in the cognitive background, framework propositions function as the intellectual foundation against which testing into the truth or falsity of further factual contents takes place, however, in Capgras delusion this intellectual background (and importantly this intellectual background alone) is in error since the proposition that My wife (child, colleague etc) has been replaced by an imposter, appears in the subject’s foundational intellectual map (Campbell 2001 pp. 96-8).

According to Campbell, the process is involved that: In a subject’s observable environment there occurs an observable stimulus which a subject recognises as object O, and which recognition (conclusion, hypothesis) by the subject that this is O is confirmed or disconfirmed by the hypothesis being tested against the subject’s framework propositions, whereupon the content that this is O or this is not O drops down into the subject’s brute motivational mechanisms where it influences (cognitively penetrates) the body’s motivational workings before bouncing back up again via contextually appropriate motivational experiences (saliences, feelings, or as Campbell says ‘perceptual affect’) and actual behaviours (thoughts, actions, language). Whereupon the subject may be heard to say: ‘That is O.’ or ‘That is not O.’ as the case may be (ibid.).

Thus, in a normal case, a subject would be in the vicinity of his wife, recognise his wife, have his system confirm that this person truly is his wife, and consequently experience a rush of welcome towards his wife, think hello wife, say ‘Hello wife’ and give his wife a hug. However, in the case of Capgras delusion, something is amiss, for within the subject’s cognitive background a framework proposition is present with the content that: My wife has been replaced by an imposter. Thus the subject would be in the vicinity of his wife, recognise his wife, but the system disconfirms his hypothesis to the end that it passes judgement: But no, this cannot be my wife! Whereupon, downstream experience and action is affected accordingly, effecting, say, a coolness or absence of feeling towards the object, coupled with the thought and possible assertion that: This person [the wife] looks like my wife, but it is not my wife; this person is an imposter. Whereupon Capgras delusion is diagnosed.

So, if we contrast the empiricist and (relationist) rationalist approaches to the problem of Capgras delusion, according to the empiricist an anomalous actual
experience comes first, which experience the subject’s brain machineries work to cognise (conceptualise, make sense of), and which cognitive processing results in the subject maintaining false conviction that: *My wife has been replaced by an imposter*, while according to Campbell’s (relationist) rationalist approach, anomalous cognitive content, i.e. the framework proposition that: *My wife has been replaced by an imposter*, comes first, and against which framework proposition enquiry into the truth or falsity of various factual hypotheses (in this case, relating to the content of the subject’s actual perceptual experience) generates anomalous conclusions which in turn affect anomalous actual experiences and behaviours.

In its favour, Campbell’s account exhibits certain strengths. Specifically, it explains the plausible-enough idea that cognitive content can affect the way in which a subject experiences the world. For example, Sally rows with Jim, and so acquires the background assumption that: *Jim is a jerk*, which will likely affect unfavourably Sally’s experiences, thoughts and conduct towards Jim for some time to come. Alternatively, Sally falls in love with Jim, and so acquires the background assumption that: *Jim is wonderful*, whereupon Sally’s experiences, thoughts and conduct towards Jim will likely take on an entirely different motivational hue. Nevertheless, there are also difficulties with Campbell’s (relationist) rationalist account.

One difficulty is that regarding Campbell’s (relationist) rationalist account of Capgras delusion, Campbell fails to credit the role that actual experience plays in events. I suggest that, in the account which Campbell provides, the role of experience is both manifold and significant. First, there is the initial experience of the object, which is an actual experience with a normal-enough content, since the subject is able to correctly identify the visual characteristics of the object as those of his wife, and without the occurrence of which actual experience no significant testing against the background propositions would take place. Second, there is the subsequent experience of affect (saliences, feelings, perceptual affect) which given the fact and role of the centrally anomalous framework proposition that: *My wife is an imposter*, continues to be an actual experience, but now it is an actual experience with oddly coloured content (anomalous saliences, feelings, perceptual affect). Third, the oddly coloured experience is not a mere sideshow, for what is now an abnormal experience of the wife feeds back into the system as conclusions are drawn re its content, and which conclusions are then tested against the subject’s intellectual background of foundational propositions, and the results of which testing drop down into the motivational underpinnings of experience.
and behaviour, whereupon the experience of said experiences and behaviours once more feeds back into the system as conclusions are drawn re its content, and which conclusions are then tested against the intellectual background of framework propositions and so on.

In other words, inherent within the dynamics of Campbell’s own explanation of Capgras delusion, actual albeit anomalously ‘weak’ experience is an additional necessary factor in play. So that in this particular case, what Campbell is proposing, could be glossed not so much as a (relationist) rationalist thesis per se, but as something more approaching a (relationist) rationalist two-factor account.

Certainly, it might be objected (or rather reiterated) that what I have identified as experience in the role of a necessary explanatory factor, and even more so as abnormal experience occupying the role, is not an important explanatory feature, but is only downstream and tangential to the actual explanatory deficit, namely the anomalous framework proposition with the erroneous content that: *My wife is an imposter*. Hence, experience is peripheral and abnormal cognitive processing is the sole significant explanatory cause. However, my reply to the counter would be that this still seems to me to be incorrect. In line with Campbell’s approach, experience and especially abnormal experience is working hard in the case of Capgras delusion. It is supplying the reasons and thereby motivating the non-virtuous epistemic feedback circle through which, and only through which, the abnormalities of reasoning indicative of Capgras delusion take hold.

However a second difficulty for the (relationist) rationalist thesis is that if is the case in Capgras delusion that experience is, as it certainly seems to be, assuming a significant explanatory role, this triggers problems regarding those cases of delusion in which experience cannot be a factor, that is to say delusions which characteristically feature zero experiential element (i.e. no experiential factor full stop) and/or a (very) strong experiential element (e.g. a strong interactive hallucination of voices or a second-head).

The reason this is problematic for the (relationist) rationalist is that if in the case of Capgras delusion actual experience plays a key role, but if in neither of the other cases of delusion can the (relationist) rationalist appeal to actual experience, then at the very least this suggests that given the dynamics of the account put forward by Campbell two different explanatory pathways are in play. The first is the conventional or ‘one-factor’ (relationist) rationalist account as per delusions which do not significantly feature
an actual experiential element. The second is a ‘two-factor’ (relationist) rationalist account as per delusions which significantly feature an actual experiential element. However, from a theoretical perspective this sort of divided approach is not the most useful, for if the test of a theory is the explanatory/predictive leverage affords across a wide range of empirical cases, here we have an instance of empirical cases determining that at this point the (relationist) rationalist theory needs to be broken and reset – plus, one might wonder, if once, then how many times more?

Again, one might hear it said that Campbell could respond easily enough by pushing everything into the cognitive, whereupon the weak experiences characteristic of Capgras delusion would be eliminated or at least re-construed in entirely cognitive terms. Yet I would argue that Campbell would likely not elect to make this move, since if the actual experiences indicative of Capgras delusion are pushed into cognitive, Capgras delusion becomes either a robust instance of a zero-experience delusion or a puny instance of a (very) strong experiential delusion. If it is the former, then Campbell faces all the problems outlined above re erotomanic delusion, but with the added difficulty that the content of Capgras delusion is even more implausible than the erotomanic content that: *S/he loves me*. That is to say, Campbell would need to provide a plausible explanation concerning not only how the framework proposition that: *My wife has been replaced by an imposter*, initially gets in there, but how it becomes established and why it or its downstream corollaries are not selected out.

Plus a more substantive concern would be that whichever way one rethinks Capgras delusion by pushing the subject’s affective responses wholly into the sphere of function/cognition a theoretical cost is incurred by the relationist. Consider Sally and Jim. The relationist view is that when Sally sees Jim, then there is something-it-is-like for her to undergo the experience. Similarly, when Sally falls in love with Jim, then one might blame it on electrical changes, chemicular changes, structural changes and/or motivational changes discharging in all directions, nevertheless, for the relationist there would still be something-it-is-like for Sally to feel love, warmth, desire, yearning etc for Jim. However, if Campbell pushes the saliences, feelings and perceptual affect characteristically associated with Capgras delusion into the same explanatory arena as that of zero-experience delusions or delusions with a strong hallucinatory component, then, unless there is something fundamentally different between experience in delusion and commonplace experience, so in the case of Sally falling in love there would be
nothing-it-is-like for her but only something-it-seems-like to know love, warmth, desire yearning etc for Jim.

Now perhaps one might be okay with this prospect, i.e. driving a wedge between the actual experience of seeing and of non-experience as regards love. Or perhaps one mightn’t be okay with the prospect – say, because one is a romantic who would argue that love is an overwhelming sensation; or because one is a pragmatist who would classify love in terms of electrical, chemical etc. changes and powerful downstream behaviours; or because one is a relationist, who would then have to deal with the sceptical concern that if Sally could be so mistaken such that what she takes to be her actual experience of love, warmth, desire etc for Jim is not in fact a genuine experience, but only seems as if it is that way, then perhaps Sally could be equally mistaken when she sees Jim. That is to say, perhaps in any given case there is never, ever something-it-is-like for Sally to experience, but only ever nothing-it-is-like but something-it-seems-like to know whatever. In other words, in shifting delusions which characteristically feature an actual weak experiential element into the wholly cognitive sphere, the (relationist) rationalist effectively undermines their own foundational theory, and although, in the process, they point us towards a nice coherent and cohesive rationalist account of delusion, this is one would think in the eyes of the relationist poor recompense.

So, finally, how does Campbell’s (relationist) rationalist account stand up to cases which are said to involve a (very) strong experiential element, which is to say hallucination, and which DSM-5 stipulates in its definition of hallucination as a clinical feature of psychosis and psychotic disorder compares to the philosophical concept of strong hallucination, i.e. ‘perception-like experiences that occur without an external stimulus. They are vivid and clear, with the full force and impact of normal perceptions, and not under voluntary control...’ (APA 2013 pp. 87-8). Moreover, as the case studies make clear, in at least some cases the strong hallucinations in question are very strong hallucinations as they characteristically or at least significantly involve a marked degree of interaction with the subject and/or the world, e.g. the subject’s ongoing discourse with hallucinatory voices or ongoing disagreement with a visual, auditory, somatic, not to say goading and belligerent hallucinatory second-head.

In short, it is fair to say that Campbell’s relationist approach to the problem of delusion has difficulty accounting for cases of delusion and delusional syndrome which significantly feature a (very) strong experiential element. Basically, the obstacles confronting it pretty much duplicate the two concerns previously discussed as regards
delusions with zero-experiential element and a weak experiential element. That is to say, at the very least, Campbell’s account of delusion faces the problem of providing a significant body of further explanation in re various features and dynamics of the theory, or, in the absence of said explanation, it faces the deep-seated consequence that, in tending the theory of delusion, the account does not support so much as subvert the relationist foundation on which it rests, to the extent that an ongoing corollary of Campbell’s account is the deeply sceptical prospect that there is never, ever something-it-is-like for one to have an actual experience, but only ever nothing-it-is-like but something-it-seems-like to judge that something is the case.

To verify this point, recall Campbell’s account of delusion as it was outlined at the start of the chapter. P1-C1 lays out details of the disjunction which is held to exist at the level of experience between the good case of perceptual experience (which necessitates a composite of actual observable objects and properties plus functional workings as the brain works to attend to said objects and properties) and the bad case of hallucination (which involves a lack of significant observable objects and properties and with it a lack of experiential content, so that all which remains is the brain working to attend as if said objects and properties were actually there). Consequently, P4-C2 confirms that in both the good and bad cases the brains workings are common in kind. Next, P7 observes that, all things being equal, when inputs (i.e. observable stimuli) into and the workings (i.e. biological, physiological and cognitive administration) of this system are okay, so its outputs (i.e. experiences, actions and beliefs) are okay as well. However, P8-C3 points to the fact that in cases of delusion outputs are not okay. Indeed, by definition delusion necessitates a subject presenting with deeply held and intractable false belief, plus in some cases of delusion (very) strong hallucinatory experiences certainly seem to occupy a significant role. However, as P9-C4 observes, in accounting for the features of this particular kind of delusion, the relationist cannot appeal to actual experience, so the root of the problem must be located in the brain’s workings and specifically in its cognitive machineries. Then P10-C5 addresses the nature of the significant anomaly of the cognitive system, and concludes that the only plausible candidate is its framework propositions. Before P12-C7 confirms that this cognitive abnormality is the sole necessary explanatory cause of the experiential, behavioural and further cognitive anomalies characteristic of delusion with which a subject presents, such as hearing voices or seeing things which aren’t there or the assertion of the
delusional conviction that, say, ‘This person [the wife] is not my wife, it is an imposter.’ or ‘This [second] head is that of my [late] wife’s gynaecologist.’

So this is Campbell’s relationist theory of delusion, and, accordingly, if one wishes to engage with the argument one cannot really jump in and criticise P1–C1. Not because one agrees with the relationist thesis, but because as with the empiricist’s appeal to the causal theory of perception and hallucination and with it commonkind experience the relationist’s argument has to start somewhere, and so the underlying relationist theory can be for now left intact. However, the downstream consequences of the theory are again a reasonable target for potential criticism. Furthermore, if the criticism stands, then the issues raised might plausibly feed back to subvert the foundations on which everything rests. As was argued above, this is what happens when Campbell’s theory of delusion is tested against delusions and delusional syndromes in which actual albeit weak (anomalous) experience is a characteristic feature which certainly seems to assume a significant explanatory role. And I further argue that the same thing happens when Campbell’s theory of delusion is tested against delusions and delusional syndromes in which (very) strong anomalous experience certainly seems to be a significant feature. And to this end I single out two experience-related concerns.

First, as regards P8 to C5 one could raise a somewhat empirical question, as to whether an anomaly in framework propositions could plausibly be said to give rise to the specific symptomologies of delusion or delusional syndromes – for instance, a (very) strong interactive hallucination of discoursing voices or visual/auditory/somatic hallucination of an obstreperous second head. If the answer is yes it can, then fair to say there is a substantial need for an explanation as regards how and why this is the case by the (relationist) rationalist. However, if the answer is no it cannot, then given that what is at point is a diagnostic criteria of psychosis and/or various psychotic disorders, this either throws into doubt the ability of the (relationist) rationalist account to comprehensively address the clinical data set, or, specifically regarding the diagnostic criterion of hallucination, it opens to door to a deeply sceptical corollary. The corollary is that since what is at issue is not merely one outlying case in which one particular subject makes one particular error of judgement re the content of his or her own conscious states, but an accepted diagnostic feature of psychosis and psychotic disorder, this means that very many patients would have to be said to be making precisely the same ‘error’ of judgement, but, given the prevalence of the ‘error’, then might it not be the case that all subjects, delusional and otherwise, could make precisely the same ‘error’
of judgement, and perhaps make it all the time. That is to say, as a matter of fact, might it not be the case that there is never, ever something-it-is-like for one but only ever something-it-seems-like to experience anything at all.

Second, as regards P5 and P6, one might raise a more logical concern, as to whether cognitive processing really can be identical in perception and hallucination in any given case, and so including or perhaps especially in a case in which an actual experience and a corresponding ersatz experience would be for the subject utterly indistinguishable by introspection alone. If the answer is no, cognitive processing is not identical across all cases, not least because, by necessity something fundamentally different is going on in a brain within which its functional workings are working in conjunction with significant objects and properties versus in a brain within which its functional workings are working around an absence of significant objects and properties, even if in both cases the outputs of those workings re experience are seemingly identical by introspection alone. However, if this is one’s answer, and putting to one side any internal problems it might raise for the underlying relationist theory, then with regard to clinical cases of delusion, we now have a situation in which a different explanation is required for delusions with zero experiential element and/or an actual weak experiential element to those with a (very) strong experiential element, namely, cases of delusion or delusional syndrome which significantly feature hallucinations which are ‘vivid and clear, with the full force and impact of normal perceptions, and not under voluntary control...’ (APA 2013 pp. 87-8). That is to say, once again a situation has arisen in which the facts of the empirical data set determine that at this particular (different) point the (relationist) rationalist account of delusion needs to be broken and reset. And, as observed above, this is not the best way of delivering an explanatory account of that particular set of empirical facts.

Conversely, if the answer is yes, cognitive processing is identical in perception and hallucination in every single case, then the door opens wide to the abovementioned sceptical concern, because if cognitive processing really is common in kind as it underlays veridical perceptual experience and hallucinations which are utterly indistinguishable by introspection alone from veridical perceptual experience, then given that what is at point is not merely a hypothetical concept but a key clinical criteria of psychosis and psychotic disorder, so, as a matter of fact, either a lot of people are drawing not just a modestly wrong conclusion but a markedly wrong conclusion as regards the content of their own conscious states – and note what is at point is not
merely a bit of commonplace conscious content which might slip past without one really attending to it, such as, whether that woman rushing past appears to be wearing glasses or not, indeed, the conscious content in question is of a sort about which anyone would not only be aware but would be given to introspect very closely, such as: Am I really hearing those voices and/or seeing my own second head?! Jeez, yes I am! Or, alternatively, might it just perhaps be the case that, as a matter of fact, a substantial number of people are not making a significant mistake as regards the content of their own mental states, that perhaps what is going on in experience and hallucination is that at least some hallucinations are utterly indistinguishable because really they are indistinguishable, not just at the level of cognitive processing but at the level of experience too. That is to say, taking it as read that we have side-lined the causal theory of perception and hallucination and are here operating within a relationist framework, then might it just possibly be the case that there is never, ever something-it-is-like for one to actually experience the sights, sounds, feel etc of, say, rain or a tree or one’s dog, but only ever something-it-seems-like to judge that this is the case?

And so the outcome of testing Campbell’s (relationist) rationalist account of delusion against cases of delusion in which a (very) strong experiential element certainly seems to assume a significant role, is that concerns arise about whether or not the account can actually meet this kind of problem. Moreover, if the theory is pushed hard so as to say that, yes, the (relationist) rationalist account can meet the problem of explaining these particular cases of delusion and delusional syndrome, then, at the very least, a great deal more explanation is required from the (relationist) rationalist concerning how and why this is so.

Meanwhile, in the absence of any supporting explanation, a consequence of pushing or ‘stressing’ the theory thus is the repeated opening of the way to the sceptical possibility that the fundamental disjunction drawn by the relationist between the subject having an actual experience, whereupon there really is something-it-is-like for the subject to see/hear/feel etc whatever, and a hallucination, whereupon there is nothing-it-is-like but only something-it-seems-like to judge that whatever is the case, is a bogus dichotomy, for the reason that in any given instance only the nothing-it-is-like but something-it-seems-like side of the division holds. This is a logical outcome, however, which opposes the central premise of the relationist thesis on which Campbell’s rationalist account of delusion rests. Still, on the upside, it is also an outcome which works to shed some light on what might possibly be a nice coherent and cohesive way of thinking about the problem of
delusion and delusional disorder. Although, when one has just undercut one’s own foundational theory, the prospect of presaging a nice alternative account of delusion would almost certainly seem like poor recompense.

All the same, as the saying goes, one man’s poison is another man’s meat. And, indeed, what works to weaken the prevailing (relationist) rationalist account of delusion really does afford a foretaste of the explanatory potency of an alternative rationalist approach. The alternative approach is a combined Dennettian and Prediction Error thesis as this is developed in chapter three and then applied to the problem of delusion in chapter four. And a strength of the account lays in its inherent flexibility. Not least because a Dennettian and Prediction Error approach manages to avoid the strict extremes which, as I have argued throughout this chapter, dog the prevailing empiricist and (relationist) rationalist accounts of delusion on account of their respective insistences that by necessity in any case of delusion anomalous experience and/or (deep) irrationality as well as significant neuropathology are or are not central explanatory factors. Indeed, as was outlined in chapter one, although a Dennettian and Prediction Error approach is correctly categorised as being a rationalist approach to delusion, is only a mild rationalist approach, in so much as unlike each one of the prevailing accounts of delusion it allows that any given case of delusion or delusional syndrome anomalous experience, and for that matter a significant neuropathology, may or may not assume a significant explanatory role.
CHAPTER 3

A DENNETTIAN PLUS PREDICTION ERROR ACCOUNT

Introduction: Moving Towards a Combined Account

In chapter one, I outlined Daniel Dennett’s intentional theory of mind and Prediction Error theory with regard to the information sorting functions of a living human brain. I observed that these are two independent theories, and so offer two dissimilar ways of talking about the world. Nevertheless, I proposed merging Dennettian theory and Prediction Error theory into a single unified thesis, that is, developing a combined Dennettian plus Prediction Error (D-PE) account of mind.

The idea which motivated this suggestion was the problem of delusion, which I identified, first, as addressing the data set of features characteristically associated with clinical cases of delusion, and, second, as specifying whether the theoretical account which best achieved this objective was an empiricist or rationalist approach. Yet, looking at the prevailing theories of delusion, at first glance, it seemed like they could each run into problems re explaining different delusions from across the data set, and this hunch was supported by a closer analysis of the prevailing theories’ respective positions in chapter two.

Thus it seems that there really is an opening for a different way of approaching the problem of delusion, and in which case a combined D-PE account seems like a reasonable option to explore. For one thing, the two theories appear to complement one another, in so much as prediction error theory offers the Dennettian model a way of thinking about what is going on at a more granular level, whereas the wide Dennettian model offers prediction error theory a way of extending its principles beyond the brain and out into the world. For another thing, it has been pointed out that a combined D-PE account can provide more explanatory flexibility than any of the prevailing theories, while at the same time slotting into the mores of the current conversation around delusion. And, lastly, given that it rests on Daniel Dennett’s intentional theory of mind (as opposed to the causal or relationist theses on which the prevailing theories rest), so it seems reasonable to think that a D-PE account really could offer a fresh approach.

Accordingly, here in chapter three, the aim is to formally merge Dennettian theory and Prediction Error theory into a combined D-PE account, before in chapter
four applying the model to the three delusions which have been used as test cases so far, namely, erotomanic delusion, Capgras delusion and perceptual delusional bicephaly.

_Tackling a merge of the accounts_

The situation so far is that there are two different theses which in principle seem to share a reasonable amount of ground. For instance, both Dennett and Prediction Error talk in terms of evolutionary principles and hierarchical systems – where ‘system’ in this sense basically means a collection of parts sharing energy and information between themselves and in so doing working as or becoming a coherent and cohesive whole, whether the ‘whole’ at point is a living human being called Sally, or the living brain behind Sally’s eyes. However, in practice linking Dennettian theory and Prediction Error theory does not present a straightforward fit. For one thing, each theory has its own language and conceptions. And for another thing, they each advance different kinds of explanation, which is a point I return to in a moment. Thus in pushing the theories together a little ingenuity has been required and applied to both sides. For instance, Dennett’s theory in some ways becomes a bit denser, whereas Prediction Error theory becomes a bit lighter. Nevertheless, other parts of the theory merge quite happily one into the other, for example the idea of Bayesian priors combines freely with that of Dennettian basic belief. Even so, given that we are not talking in terms of an exact equivalence between the two theses, but in terms of an interpretation of how they may fit, this suggests that it is likely that the two accounts could be combined in a number of related but dissimilar ways, two of which I mention in a moment. However, for this reason, the way in which I approach the merging of the theories is the way of combining Dennettian theory and Prediction Error theory which makes the most sense to me.

_So how do I go about it?_

The position is that the principles of Prediction Error theory need to be slotted into the framework of Dennettian theory, as both the principles and the framework were outlined in chapter one, so that the dynamics of prediction error reporting (in theory) offer the Dennettian model an additional finer-grained layer of description about what is going on.

Consider Dennett’s theory. It is an account about how a fundamentally rational human being (system) with a very, very long history of having evolved through many
different stages of development is motivated by need to satisfy its basic requirements as it by necessity negotiates its environment and in so doing survives to tell the tale (or not) (Dennett 1987 [1989a], 1991 [1993], 2017). As Dennett observes, in principle there are various stances one could adopt to supply this account (Dennett 1987 [1989b]). For instance, one could adopt a physical stance and deliver an explanation from the level physics and chemistry; a design stance to explain from the level biology and its systems, structures and functions; or an intentional stance, which is the level from which Dennett elects to explain things in terms of the patterns of observable behaviours by way of which the system (subject) interacts with relevant criteria in its world, and which (patterns of behaviour) at the level of the (minded, intelligent) subject are represented by the culturally accepted mentalistic terms of folk psychology, e.g. needing desiring, believing, judging etc. (Dennett 1987 [1989a, 1989b]). Thus in providing an explanation from the intentional level (i.e. by adopting an intentional stance), on the surface at least, there is no obligation to think in detail and at length about the particulars of physics, chemistry and biology, but one can legitimately talk about, and so provide a swift coherent explanation in terms of, Jones hankering for beans, or Sally loving Jim.

Now think of prediction error theory. Prediction error theory, at least as it is developed and applied by Friston and others (e.g. Friston et al. 2006) is a theory about learning and perceptual inference in a biological system (specifically, a brain) whereby its emergent behaviours are explained in terms of a ‘probability distribution’ between principles encoded in the ‘system’s configuration’ and the ‘environmental quantities’ which act on that system, and where the difference between expectation and fact is measured in terms of ‘free energy’ (i.e. prediction error), and in which ‘the system can minimise free energy by changing its configuration to affect the way it samples the environment or change the distribution it encodes’ (p. 70). In other words, prediction error theory is a design level explanation, and so not really the right sort of instrument to be talking about hankering for beans or being in love.

Thus an immediate difficulty of merging the two theses is that while Dennettian theory and Prediction Error theory appear to be talking about similar things in the process they are generating two different kinds of explanation. So if the aim is to provide an account within the framework of Dennettian intentionality, then the question is how and where does the design level mechanics of prediction error fit in? As I observed a moment ago, there are likely to be different ways that one could approach this question. Let me consider two.
The first way, which is perhaps the more conservative approach, potentially gives more space to prediction error theory as it is talked about by Friston and others, i.e. as a separate theory. We can call it ‘the flight control at midnight’ D-PE model. And it portrays a Dennettian system, at the hub of which the brain sits in utter darkness at the top of a control tower, where it is continually engaging with huge banks of electrical data, and in so doing directing with a high level of accuracy a machine which it can’t see, can’t hear, can’t feel etc. If one favours this line of explanation, then perhaps something like a dual account could be given, first the Dennettian side, which is then paralleled and supplemented by a description of the predictive goings on inside the control centre of the brain. It is a feasible approach. Though perhaps not the ideal option if the aim is to present a unified philosophical thesis. Plus we again run into the question of prediction error’s suitability for talking about intentional states, because ‘probability distributions’ just don’t seem to hack it when we are thinking about personal concerns like Jones’ hankerings or Sally’s love.

The second way, which is the approach I use here, literally merges the two accounts. And since the outcome neatly parallels Aldous Huxley’s metaphor of the brain as a ‘reducing valve’ (Huxley 1954 [2004]), so we can call it ‘the system as a reducing valve’ D-PE model. The principle behind the account is that in any situation there is a lot of information moving around. This is noise. And noise is not just noise, it is a cacophony, an incredible incomprehensible body of shifting competing patterns (Dennett 1991, 1991/1998 [1998b]). Where patterns are regularities, which is to say parcels of data with potentially identifiable, and in some cases familiar, edges (Dennett 1991). For instance, a daisy flower, a melody, and the word ‘BOAT’, are all familiar regularities, i.e. informational patterns, within an overall body of informational noise. Yet what this means is that by necessity a biological (i.e. living) organism, if it is to remain a living organism, needs to select meaningful patterns (e.g. regularities pertaining to criteria involved in respiration, nutrition etc.) and only meaningful patterns from out of the surrounding noise. That is to say, if it needs an O$_2$ molecule there is no point in it selecting a daisy flower. And so the system must trim down the totality of information available to it to only those informational patterns which are going to facilitate its getting by. And according to the system as a reducing valve D-PE model as I present it here, it is the job of the brain not so much to do as to assume a role in this massive task – which is not quite the same as Huxley’s original conception, but it remains a close enough comparison nonetheless.
To make biological survival possible Mind at Large [i.e. the totality of available information] has to be funnelled through the reducing valve of the brain and nervous system. What comes out the other end is a measly trickle of the kind of consciousness which will help us stay alive on the surface of this particular planet. (Huxley 1954 [2004] p. 11.)

So, in contrast with the D-PE flight control at midnight model, the D-PE system as a reducing valve model as I approach it conceives the (biological) machine as acting and interacting with itself and with its world – which all things being equal includes other (biological) machines which are also acting and interacting – via a huge compression of information on account of which the deluge of detailed information transmitted by the electrical data bank is disregarded in favour of coarser-grained informational patterns, which are likelihoods, fictions, stories, i.e. compressed retellings of its circumstances, but which are fast and on the whole accurate enough to tell apart O₂ from a daisy flower, and are just the right sort of thing for talking about human needs, desires, and states such as love.

Take an example to illustrate. Consider an environment, it contains blue skies, green fields, and in the distance rolling waves. And into this setting place two people, say, Sally and Jim. Now if we start thinking in terms of information, it is immediately apparent that there’s an overwhelming jumble of it! There is information in the brain, but a Dennettian system is wider than the brain. So in addition, there is all of the information circulating throughout the biological body, and all of its systems, organs, and functions, including that of the skin, at which point a Dennettian system carries on outwards into the body’s intimate, immediate and wider surroundings, which includes its interpersonal setting (so now we have two brains and two bodies to consider), and then outwards further into the cultural setting (which includes among other things its own panoply of goods and bards, dos and donts). Then factor in that throughout this entire context information is continually moving around, that nothing ever stays still. Yet in this informationally incomprehensible context, when Sally looks at Jim she, or rather her system, needs to act fast so as to reduce, i.e. extract from the whole, those patterns which are informationally meaningful (important, valuable) at that particular moment, while simultaneously selecting-out the rest. And where what needs to be selected will be
substantively different if Sally’s intentional state when she looks at Jim needs to be *I love you* as opposed to *you jerk*.

Again there may be different ideas concerning how in this sort of situation relevant information is selected-out of the body of noise. Yet in line with the system as reducing valve D-PE model, I propose that the process of selection is neither random and nor does it happen in one step. Instead, information is selected-in and so reduced in parallel across the wide-ranging system in many, many connected stages, and where each step or stage functions in a fashion which is amenable to description in terms of prediction error reporting. So that a living human system simultaneously across many hierarchical levels of function is continually predicting and testing the content of new informational content on the basis of the old information it already holds. Thereby the system effectively and progressively ‘talks to itself’ so as to narrow its options and in so doing is drafting a rich and meaningful story which, perhaps, eventually, if circumstances call for it, the subject, e.g. Sally herself, can use to interact with her world.

*Developing the combined account*

So this is the beginning of a wide and deeply interwoven way of merging Dennettian theory and Prediction Error theory. But I need a way of talking about the idea in more detail, i.e. developing a model account. In the past I have tried approaching this in the abstract, with mixed results. So in view of this, and certainly in keeping with the spirit of the thesis, I choose to use a case study – in this case an entirely *normal* case study, in contrast to the *abnormal* case studies involving delusion outlined in chapter one and returned to in chapter four. And I use the normal case study as a scaffold to build the D-PE model around.

As with the case study of erotomanic delusion involving Patsy, the case study I use here does not describe an actual series of events. It is an exemplar, a fiction, but a good fiction nonetheless, in so much as it describes a series of events which provide sufficient detail to talk about the problem in a meaningful way. The case is a borrowed and expanded retelling of one of Dennett’s own illustrative examples from the book *Brainsstorms* (Dennett 1978 [1981a]), where it appears in the chapter *How to Change Your Mind* (Dennett 1978 [1981b]). The gist of my approach is that it considers from the perspective of D-PE theory the changing situation (inside and outside) of a subject, Dennett, as he approaches the potential purchase of a boat, the sloop *Molly B*, first
gradually making up his mind to buy it, before (perhaps) changing his mind and pulling out of the deal.

My rationale for using this particular case study is that if delusion is by definition a case in which a subject makes up his or her mind that something is the case, and then (oddly, abnormally) holds onto the conviction despite all reflection, reasoning and evidences to the contrary, then an explanatory model re a parallel situation in which a subject makes up his mind, and then (unexceptionally, normally) changes his mind in light of reflection, and/or reasoning, and/or evidences, could provide a reasonable base from which to think about what might be going on in the abnormal case as well.

Thus with the case study in place, I start to develop a combined D-PE account. And to do so I present the explanation in several steps.

The first step is to present the normal case study of Dennett and the Molly B.

The second step is to think through the principle of a wide Dennettian system coming together with the rules of prediction error reporting so as to deliver a unified D-PE account. A crucial step in this regard is that the Dennettian system needs to be re-represented as a formal and connected hierarchy of stages in accordance with the requirements of prediction error theory, before detailing the process by which prediction error reporting controls the movement of information between and across all levels.

The third step moves beyond the conception of a system per se, and in so doing takes the combined D-PE model and applies it at the level of the subject, i.e. Dennett, as he makes up his mind to purchase the sloop Molly B, before quite understandably changing his mind and reneging on the deal.

The fourth step stays with the subject, Dennett, to think a little bit more closely about the account from the perspective of his perceptual experiences, and specifically what the combined D-PE model can tell us about what happens when Dennett first sees the Molly B.

The outcome of developing and applying the D-PE model in this way is that if all goes to plan by the end of chapter three I should have a reasonably sound and comprehensive Dennettian plus Prediction Error model which I can take forwards to test against the three cases of delusion in chapter four.

Finally, as I work through the chapter, I aim to address as many points and queries as I can given the time and space available, and for bringing many of these ideas to my attention during a presentation I gave on Dennett’s essay *How to Change your Mind.*
I would like to thank Professor Tom Stoneham and fellow students in the Work in Progress group at the University of York.

The Normal Case Study of Dennett and the *Molly B*

The entirely unremarkable series of events surrounding Dennett’s wanting and potential purchase of the sloop *Molly B* is as follows:

1) There is a normal (typical, healthy) subject, Dennett, who develops a hankering for a boat.

2) The hankering is nothing specific. Just a vague boatish urge. Still, it’s enough to get Dennett collecting apposite facts, say via TV, magazines and books. And the upshot of all this fact gathering is that Dennett starts to narrow his options, so soon it is not just a boat he’s desiring, but a sloop with properties a-b-c-q-r-s.

3) Equipped with his list of desirable properties, Dennett visits the boatyard, where he is met on the jetty by salesman Bob. Dennett tells Bob his requirements: ‘A sloop with a-b-c-q-r-s.’ Whereupon Bob, as salesmen are apt to, says: ‘That’s, great! I have three boats which meet all your requirements. They are *Flight*, *Unicorn* and the *Molly B*.’

4) First they visit *Flight*. However, the instant Dennett sees that *Flight* is purple, he seamlessly amends his list of desirable properties to include that of being not-p!

5) Second they visit *Unicorn* and *Molly B*. These are two tantamount identical sloops moored side-by-side on the jetty. After a moment of consideration Dennett gives preference to the *Molly B*.

6) Bob senses a sale and leaps into action. He knows that: (a) you need to strike while the iron’s hot, and (b) there is a world of difference between Dennett preferring and Dennett buying a boat. He has to get Dennett to commit to the transaction. ‘It’s one of a kind,’ Bob lies. ‘Reduced only this morning from $$$$$ to $$$$. It won’t be available for long,’ he says, perhaps truly enough.
7) So what happens next?
   a. Does the thought of parting with $$$ cause Dennett’s desire to evaporate, whereupon he walks away?
   b. Does Dennett think of blue sky, white sails and sea spray, setting the issue of $$$ aside, and so okays the deal and sails off into the sunset.
   c. Does Dennett commit to buying the Molly B, but then later, perhaps at home after a bit of reflection, change his mind and decide that he doesn’t really want the boat after all, whereupon he phones Bob and reneges on the deal?

There is nothing irrational about any of these outcomes. Indeed, any one would make a plausible conclusion to events. However, the task now is to explain what is going on throughout the case study with regard to Dennett’s desires, thoughts, actions and decisions from within the framework of a combined D-PE account.

**A Dennettian plus Prediction Error Account**

**A Combined Dennettian plus Prediction Error Framework**

In chapter one, I introduced the core principles of a Dennettian system and those of Prediction Error theory. So in this section the aim is to take both sets of concepts and combine them into a single theoretical model. Namely, a Dennettian plus Prediction Error (D-PE) approach. The way I tackle the merging of the theories is, in line with an essential requirement of prediction error theory, re-representing the Dennettian system formally in terms of a connected hierarchy, and then clarifying how the principles of prediction error reporting (might plausibly) function to control the movement of energy and information throughout.

So the first thing is to clarify what a system in the Dennettian sense involves. As previously noted, this was introduced in chapter one, but now I need to reintroduce the concepts and say something more about them, especially as regards the system’s reasons and motivations, in light of the work which I am asking them to do here – and importantly which I will be asking them to do when I apply the model to talk about delusion in chapter four. Accordingly, I start by thinking about the system from the perspective of the living human subject, who for explanatory purposes – and just for the moment – I divide two ways, so as to talk about the human system (i.e. a living human
being) in terms of its ‘bottom-end’ and ‘top-end’ features and capacities, before I expand the system outwards to consider how the human subject fits in to a wider environmental context.

At the bottom-end of the human system

The bottom-end of the living human system is basically its brute biology. That is to say, the systems, organs and functions (i.e. biological machineries) which comprise an individual body, and on account of which the system is a living system, and which parts individually and collectively work together with the sole ‘hardwired’ (structural, mechanical) objective of keeping the system (living body) working as it ought to work, by design. Design in this sense means evolutionary design. And specifically re the functions and capacities of bottom-end biology we are thinking about the long slow influences of evolutionary changes through the dissemination between individual organisms of basic biological materials, i.e. genes (e.g. Dennett 1991 [1993], 2017).

In approaching the system via its biological machineries the position one adopts can be as granular as one’s specialism or requirements demand. However, given that I am not aiming for a design level explanation per se, but an explanation from a Dennettian intentional stance, so the level at which I approach biology can (indeed, must) be pretty coarse-grained. Accordingly, I refer loosely to the systems, organs and utilities of the living body, with a little more attention being given to the nervous system – literally at the top of which sits the brain controlling operations. And of special interest is the functional role played by the nervous system in organising responses re the occurrence of certain highly typical criteria in the system’s (i.e. body’s, organ’s) environment, and which (criteria) we can gloss as falling into two basic camps of harms or boons.

The gist of the idea is that in the roots of its biology for a living system to remain a living system, i.e. preserve its integrity, it has certain deep requirements which need to be continually maintained (cf. Dennett 1987 [1989b] p. 20, 1991 [1993] pp. 176-82). Essentially, satisfying these requirements supports the replication of (parts of) the system, where replication in this sense means something like renewal or regeneration, but also reproduction via sexual behaviours through which biological (i.e. genetic) material is shared and (quite literally) renewed (Dennett 1991 [1993] pp. 173-6, 2017 p. 47). A consequence of this is that at any given time there is an incredible amount of brute motivational energy driving any living system from the bottom upwards towards
certain important ends. Another consequence is that even if for whatever reason a particular human subject does not engage in sexual activity, the disposition towards interpersonal contact and intimacy is deeply ingrained, i.e. ‘hardwired’, into the fabric of his or her biological design.

Thus thinking about the system from the perspective of a particular human subject, it has a fundamental requirement to avoid things which compromise its integrity, such as actual damage, and towards things which sustain it, such as oxygen, movement, nutrition and interpersonal contact. We can call these its (the system’s, and by extension the subject’s) basic needs.

So thinking back to the role of the nervous system and a body’s brute responses to typical harms and boons. By design, deep inside its workings, the living human system mobilises with lightening rapidity certain patterns of behaviour which in the past have tended to get it away from criteria which have tended to cause harm, i.e. undermine its integrity – this is a negative motivational response – or get it close to criteria which have tended to sustain its integrity, i.e. boons – this is a positive motivational response – while roundly ignoring everything else (Dennett 1991 [1993] pp. 173-82; Bromberg-Martin et al. 2010). So, in this way, working precisely as it ought to work so as to sustain its needs by means of a continuous progression of avoiding and engaging and ignoring, the biological system (re its parts and the subject) negotiates the changing circumstances if its world moment by moment, day after day.

So looking a little more closely at the nature of negative and positive motivational responses. On the one hand, negative aversive responses re a criterion (or rather, a pattern of information in connection to a criterion) judged (by the brain) as (likely) detrimental trigger particular patterns of (bodily) behaviour designed to mobilise the system away from the (adjudged) danger in a manner which has proved itself reliable enough over time, and perhaps the most well-known example of which is the Fright, Fight, Flight response re a (adjudged) harm. On the other hand, positive responses re a sought after criterion (or rather, pattern of information) judged (by the brain) as (likely) desirable, trigger patterns of behaviour designed to mobilise the system towards engagement with a (adjudged) boon, for instance towards Food, Freedom of Movement, or the other F which Dennett tactfully refers to as Mate (Dennett 1991 [1993] p. 188). Yet to achieve these ends, the system needs to do several key things. First, it needs to recognise a pattern of information as important. I say more about this shortly when I talk about prediction error. Second, after deeming the pattern important,
it needs to assign value, either negative (avoid) or positive (engage). Then third, the brain needs to inform the relevant bits of system about what is going on (and with this, what to do next) by releasing a torrent of related electrical and chemicular changes into its wider surroundings. And the effect of this on the system is immense. For instance, some bodily systems can literally close down and others speed up as blood is literally diverted away from some organs and towards others as the entire body adapts so as to attend to the (adjudged) important criterion at point.

However, as powerful as these changes can be, the important thing to remember is that they are spontaneous functional responses, and so not the sort of thing about which any subject (without the use of measurements obtained via scientific instruments) could ever be aware. For example, deeper and faster than any person can think a brain judges danger or desire, whereupon, and without the subject ever being aware of it, blood is diverted towards his or her heart and the pupils of his or her eyes dilate. However, some upstream changes may have downstream effects which are (potentially) observable. For instance, after the event, the subject might reflect that he crossed the room swiftly; that his heart is beating wildly; that his attention is fixed on $x$; that he is perspiring or shaking; and that he is feeling terror, or anger, or agitation, or excitement, or yearning, or lust.

Possibly, at this point, it may be queried that I risk muddling the idea of design. On the one hand, I am appealing to evolutionary design, i.e. natural selection, the consequences of which are the upshot of random (blind) variations. On the other hand, I am (or will be) appealing to the principle of design to say that the system selects, or chooses, the patterns which it expects to find. In response, I submit that given the way in which they are being used here there is no real tension between the two conceptions. First, imagine that at base, deep in the physical roots of being, blind causality is all there is. There is no intention, no mind, there are just bits, bumping and banging around in the darkness, whereupon the upshots of interactions might fail, or might work, but in any case these are random connections nonetheless. Now imagine that through a long enough series of random connections the consequences of some of the luckier upshots start working together as a (rudimentary) system, whereupon the dynamic is no longer quite merely bump and bang, but exhibits the rudimentary characteristics of intention, in that to remain a system it (the system) needs to address certain basic requirements, which it does by anticipating what it expects to find. However, the moment that a system, no matter how rudimentary, starts exerting even the most undeveloped
preferences upon its surrounding environment, then on that particular stage selection is no longer merely a blind bottom-up game. It is an ‘intentional’ game as well. Plus one might think that as systems evolve, so the capacity to interact or anticipate or choose might evolve in tandem, so that gradually top-down influences could start to exert a more obvious role (cf. Dennett 1984 [2002b], 2003 [2004]). Nevertheless, even at the lofty level of the personality, no matter how crafty or clever or in command of its own destiny any particular human system might like to flatter itself that it is, each personal selection that that human being makes is still at base a roll of the dice into the dark.

Finally, one might ask, why is the idea of needs, brute motivation and associated bodily changes important to the account? The first answer is simply that it is a key part of the theoretical framework. Namely, it explains the basic biological foundations and spontaneous motivational responses of a Dennettian system, while also providing some explanation of the energies sustaining the prediction error reporting which (I suggest) is happening throughout it. However, a second answer is that, more than its being just a bit of theory, we can actually observe these responses (or rather, some of their downstream effects) at work. For instance, in the normal case study it is the same old motivational machineries re avoid and grab which drive Dennett’s genteel hankering for a boat. I say more about this soon. Furthermore, in the case studies of delusion, I argue that a significant part of the explanation lays in precisely these responses driving the various subjects (systems) from deep in their workings – and potentially driving them (very) hard. Indeed, both Coltheart and Campbell, in their respective accounts of Capgras delusion, appeal to this factor when they each allocate an explanatory role to the effects of ‘autonomic’ changes on the subject (Coltheart et al. 2011 pp. 283-5; Campbell 2001 p. 92).

At the top-end of the human system

The top-end of the human system houses its higher-order cognitive capacities. That is to say, a particular subject’s capability and tendency towards thinking, reflecting, analysing, abstracting etc. about the circumstances of its (or indeed any possible) world. Hence in contrast to the biologically defined highly typical nature of the system at the level of its bottom-end workings, at the top-end things are much more plastic or flexible, in so much as there is scope for ‘softwired’ learning and so variation between the thoughts, reflections etc. which individual persons may entertain (e.g. Dennett 1991 [1993] pp. 182-93, 193 etc., 2017 pp. 99, 159-60, 303).
Importantly, however, within the Dennettian framework, cognition is not viewed as distinct in type from a body’s basic biological functions, nor indeed from the experiences a subject has. Rather, Dennett’s proposal is that bottom and top are two complimentary ends of one massively connected system which has been laid down, layer upon layer, by dint of evolutionary dynamics over time (e.g. Dennett 1991 [1993], 1984 [2002b], 2003 [2004], 2017). In other words, the newer ‘flasher’ cognitive capacities of the modern human system are not separate from but remain connected to the older ‘housekeeping’ functional bits of the human system. The key point of this being that via this connection the old hardwired response mechanisms persist in exerting a bottom-up selective effect on a subject’s higher capabilities, i.e. experiences and thoughts. In Dennett’s words, the old machineries continue to drive matters and in so doing influence a ‘vestigial positive or negative editorial spin’ all the way up (Dennett 1991 [1993] pp. 176-82). Yet the reverse holds too. That is to say, experiences and thoughts have a top-down influence which potentially reaches all the way down into the roots of biology. This is easy enough to demonstrate, just look at or think about something which you find markedly unpleasant or pleasing and observe the effect that it has on your body.

Yet although brute motivation exerts an influence on the contents of a person’s thoughts and intellectualisations all the way up to the very top-end, in the course of its push into cognitive processing motivational impetus typically loses a lot of its urgency and brute oomph. As a general rule, we can say that from the initial picking out of a need satiating stimulus to the actual behavioural effects re that stimulus being performed, the more layers of processing which are involved, so the slower and more attenuated the response will be. Think of the normal case study. The subject, Dennett, has a basic need, say, for Freedom of Movement, in connection to which his system frees up a quantity of positive motivation in the direction of a boat. However, the result of this is not a mindless rapacious grab at the nearest available dinghy. Instead, given the time and the space to process the relevant information, there is a working through of options, i.e. a sourcing and sorting of pertinent data through many competing rounds of (basic) belief and judgement, every round of which uses energy, so that in due course what breaks through into Dennett’s overt behaviours may be something like an unspecified disquiet or, as the case study stipulates, an indefinite boatish hankering on which Dennett may subsequently reflect.
Thus an upshot of the brute biological system developing and using a capacity for judicious and mitigated informational processing is that the higher-order system, indeed the higher-order subject, gains room for manoeuvre in relation to circumstances of its world (e.g. Dennett 1984 [2002b], 2003 [2004]). That is to say, an intelligent system (subject) is no longer bound to acting mindlessly re a narrow set of typical stimuli by way of a narrow set of typical responses. Instead, with increasing cognitive capacity there comes an increasing capability for thought, reflection and agency (volition, choice), and all of which confers on the human system the (possible) option of satisfying its basic requirements in (potentially) some not so basic ways. For instance, the system’s basic needs for oxygen, food, movement, interpersonal interaction etc. might now be satisfied via a subject’s more mindful desire for perhaps relatively novel criteria, such as, fresh air, an omelette, romance, or as described in the case study, a boat (e.g. Dennett 1987 [1989b, 1989c]).

Widening the system to include its multiple environments

The question now must be How and Why does a fundamentally biological system decide that it needs, of all things, a boat? I consider two answers.

The first answer is in its environment. Accordingly, this is the point when we extend the system outwards from the narrow confines of a living human body and into said body’s wider world (e.g. Dennett 1987 [1989c], 1991 [1993] pp. 182-7). The reason for making the move is that motivation and information do not stop at the skin. Indeed, there are vast amounts of both in any typical environment, as the earlier example of Sally and Jim showed.

So thinking back to Sally and Jim, we see that their environment is in part a matter of natural phenomena, such as blue sky, green fields and breaking waves. We might call this the subject’s physical environment. However, a subject’s environment also comprises many other living organisms, such as trees, birds, pet dogs and importantly other human beings, such as the subject’s family, friends and associates, plus the countless interpersonal connections these entail. Indeed, Jim’s environment contains Sally, and Sally’s environment contains Jim, whom we know Sally either loves or considers a jerk – or then again, thinking back to the feckless, faithless and abusive lovers from chapter two, who Sally may love and consider a jerk. So we may call this a subject’s interpersonal environment. However, then we can push out even further into the wider social and cultural environments within which a subject lives, and all the
pressures and connections and struggles these contain. I say more about this in a moment. But first, two things that I want to point out about the environment in all of the above senses are that: one, individually and collectively human beings interact with and in so doing learn from and about their particular environments (Dennett 1991 [1993], 2017); and two, in any normal situation a subject’s environment contains an immeasurable amount of stuff to (potentially) learn about. That is, in any typical situation, an environment is stuffed beyond capacity with informational noise, where noise in this sense isn’t mere noise, it is noise qua a totality of patterns, and as such noise is the medium in which (all) patterns survive and move around (Dennett 1991, 1991/1998 [1998b]).

Yet within the body of noise not all patterns are equal. For instance, some patterns are important, e.g. language, many patterns of which support our interpersonal and intellectual needs (e.g. Dennett 1991 [1993] pp. 193-9, 227-52). Many patterns are pervasive, e.g. the habitual goods and bads and dos and donts of a given population (pp. 199-208). And many more patterns are (ostensibly) irrelevant, e.g. a beer can hat (ibid.). Nevertheless, according to Dennett (1991 [1993], 2017) following Richard Dawkins (Dawkins 1976 [2016]), one way in which all of these patterns are equal is that each one of them is an instance of a meme (e.g. Dennett 1991 [1993] pp. 199-208; Dawkins 1976 [2016] pp. 245-260). A meme is a replicable pattern of communal behaviour. And as with genes through which patterns of biological ‘behaviour’ replicate by way of evolutionary principles via narrowly interpersonal interactions, so memes replicate through populations via potentially exponential interpersonal interactions. Accordingly, unlike genes, memes can spread with incredible speed, since in a given case a single subject (system) may interact communally with many others (subjects, systems), and some or all of whom may judge the interaction as important, and so come to adopt the behaviours as their own as they subsequently interact with many others, and so on. So, for this reason, certain rules, values, customs, habits, words, tics etc. proliferate. And so we can easily conceive a situation in which Dennett (whose system, deep in its fabric, is driven by a need for Freedom of Movement) sees boats on the ocean, overhears a conversation about boats, sees boats on the TV, so that quite subtly patterns of boatish information start being dragged inwards, i.e. become increasingly meaningful to him.

It might be queried, if this is the case, then why doesn’t everyone want a boat? And the answer is that everyone might, but the more likely state of affairs is that not everyone has a background identical to Dennett’s. For instance, Alice might never have
visited the ocean, or her uncle might have drowned at sea, or she might prefer horseback riding, and consequently Alice may be less inclined than Dennett to covet a boat.

Again it might be queried, okay, but why is this relevant to the account? And as before the first response is that because it is a key part of Dennettian theory. However, more specifically, looking ahead to the abnormal cases, such as Patsy’s case of erotomanic delusion, these points about the environment, and norms, and noise, and the pervasive pressure that they collectively exert on a subject and thereby on her desires and thoughts and other behaviours are I propose a key explanatory point.

So now returning to the question of How and Why a fundamentally biological system decides that it needs, of all things, a boat. The second answer is tied to the cognitive capacities of the human mind. That is to say, a consequence of the human system’s increasing cognitive development is its capacity to employ information as it needs to, long after an actual stimulus has passed from the immediate surroundings (e.g. Dennett 1991 [1993] pp. 1987-193, 2017 pp. 347-8). In other words, unlike an organism such as a sunflower whose behaviours are strictly regulated by the actual presence (or not) in its immediate physical environment of need satiating criteria such as water, sunlight etc., a human being can retain, recover, associate, romanticise or even fantasise about the conception of sunlight, so that even in the depths of a Yorkshire winter one can expect or moodle – and indeed moodle dismally, or contentedly, or hungrily, or excitedly, or elatedly – about sunshine, or dog days, or strawberry ice-cream sundaes, or riding a solar flare to the moon.

And, in a sense, much of this (cognitive) activity transpires in yet another sort of environment, which we might think of as an ‘unreal’ (intellectual, logical, counterfactual) environment, or indeed as an environment of virtual presences in virtual space (Dennett 1991 [1993] pp. 288, 291, 360-2).

Accordingly, in the right set of circumstances, if Dennett’s system were to be motivated (prompted) in the right way, then various boat-related concepts which are (potentially) accessible may stimulate via top-down pressures deeper motivational changes re boats or boat-related concepts, and do so in the absence of any actual boat.

Furthermore, with regard to delusion, I suggest that the capacity of a human system to recall, retain and redraft information re significant criteria in the actual absence of said criteria, as well as, in certain circumstances, there being a susceptibility for a slide into increasingly ‘unreal’ environments is a central concern.
Representing this system as a formal hierarchy

The task now is to merge the idea of a wide, deep, motivated Dennettian system with prediction error theory. And to do so the first step must be to re-represent the Dennettian system as a formal hierarchy. This is because, within the prediction error model, the system (brain) conceived as a (massively) connected hierarchy is the essential arrangement by way of which information is shared between the various levels of processing and in so doing the system cultivates a rich dynamic conception of its world.

The use of hierarchical models enables the brain to construct prior expectations in a dynamic and context-sensitive fashion. (Friston et al. 2006 p. 70.)

The hierarchical aspect is important because it allows the brain to learn its own priors and, implicitly, the intrinsic causal structure generating sensory data. (p. 70.)

The notion of a hierarchy rests upon the distinction between forward and backward connections. (p. 79.)

Thus to re-represent the Dennettian system as a connected hierarchy of ascending/descending functional levels, I propose reducing it to six connected stages of informational processing. In descending order, these stages are: Observable Behaviour; Judgement; Belief; Motivation; Brute Biology; and the surrounding Environment (see below). Concerning the concepts of ‘judgements,’ and Dennettian ‘belief’, these were introduced in chapter one, and I return to them shortly. However, the crucial point now is that while the hierarchy is linear, it is also a loop, in that at its top-end observable behaviour doubles back and feeds into the environment at the bottom-end by way of effecting patterns of change within it, and which patterns are thereby incorporated into the body of environmental noise – ‘which then provides the targets for the next round of interactions’ (Dennett 2007 p. 263). And so the cycle continues.

THE CONNECTED HIERARCHY OF DENNETTIAN INFORMATIONAL STAGES

OBSERVABLE BEHAVIOUR
- spontaneous or volitional
- motor or linguistic (internal (thought), external (speech, writing))
- personal experience

The information in patterns of behaviour feeds into the environment, and so is incorporated into the body of noise, whereupon the patterns are potentially observable, but not necessarily observed.

**JUDGEMENT**

- instances of context fixation in the brain
- the effect of the brain’s opting for a particular outcome

Locatable in (neural) space and time, but scattered, fragmented, and at various stages of drafting, yet together realising something like a narrative stream. Namely, the hidden story which the brain is continuously telling itself about its world, which is under ever-changing revision, and which if it is prompted in the right way releases pertinent patterns of downstream changes, some of which may lead to apposite patterns of observable behaviour.

**BELIEF**

- patterns of basic informational contents disseminated across the system
- behaviour disposing organs
- continually shifting and evolving as circumstances change, and new content is added, and old content is amended or eliminated, and as the living system obsessively functions to reduce (narrow) its options (possibilities) re what is going on in its world

The belief system constitutes a massively unstable and competitive environment of losers and winners in which failing contents may be redrafted or eliminated, and successful contents may remain in circulation or achieve a (transient) state of relative stability as a brain judges (i.e. opts for a particular outcome, or places a bet on the truth).
**Motivation**

- motivational (affective) states, e.g. desire, aversion, ennui
- brute motivational responses, i.e. grab, withdraw, ignore

In a human subject motivation supports a hierarchy of stages which reflect the evolving or developing capacities (or intelligences) of the system, and where each advancing stage (potentially) affords more and more *room for manoeuvre* between the initial occurrence of the stimulus in the environment and the system’s (subject’s) concluding response.

**Brute Biology**

- the systems, functions and utilities of a living body
- nervous system, and specifically the brain
- motivational control mechanisms

**Environment**

- the physical environment
- the interpersonal environment
- social and cultural environments

In addition, there are ‘unreal’ (intellectual, logical, counterfactual, virtual) environments, within or about which an intelligent subject can expect or predict or create. For instance, the theoretical environment over the horizon, or in which the sun shines tomorrow; or the fantastical environment where a solar flare can be ridden to the moon.

Certainly this representation is a (massive) simplification of the (massive) complex of information, interplay and changes which are continually happening with regard to a living human being as he or she negotiates the world. Nevertheless, it serves well enough to illustrate the different stages of processing as per the Dennettian conception, as well as the up and down pathways of energy and information exchange as per prediction error theory.

Specifically, it provides a way of visualising how across various levels of processing, level-by-level, informational content might ascend *all the way up* and/or
descend all the way down. For instance, re the bottom-up route, imagine that in a physical environment there is a tree and a man; and from the man’s (or rather, his system’s) perspective, informational patterns in connection to the tree are witnessed and selected as important; whereupon information in those patterns is incorporated into the melee of this particular subject’s basic belief set; at which point, the patterns fracture as the information they contain disseminates within various specialist regions where it vies alongside competing contents; until (perhaps), and in some configuration, the contents gain a degree of stability, and thus some transient purchase in the brain’s own narrative; so that if prompted – at the right time and in the right way – the man himself might experience a tree, and/or in thought or in speech articulate the phrase, an oak.

Conversely, in a top-down direction, information can feed or fall down towards the body’s brute biology and/or the body of noise. I say more about this re prediction error processing in a moment. Plus, in some circumstances, the movement of information will be neither up nor down, but outwards or laterally across a particular level. Again when I talk about prediction error, and specifically as regards the idea of information gaining lateral purchase at a particular stage of processing, this will be an important idea.

A final point is that in no circumstances is there any guarantee that any pattern in the environment will successfully navigate the passage upwards into observable behaviour, including the behaviours of experience and thought. Likewise, there is no guarantee that content which starts at the top will successfully navigate its way downwards. For instance, think of all the good intentions which go astray.

**Prediction error reporting**

So with the Dennettian system re-represented as a connected (circular, or cyclical) hierarchy, the question is how does prediction error reporting fit in? And I suggest that, with all of the above conditions in place, the principles of prediction error reporting apply themselves in a rather straightforward way.

To explain why I think this is the case, first consider that here we have a Dennettian inspired hierarchy. It is cyclical, for the reason that the level at the top feeds back into the level at the bottom. As such, it has a number of levels or stages, each one of which is located between the level below and the level above.

Now think back to chapter one, where I introduced and described the principles of prediction error reporting. The idea is that at any particular level of processing,
specialist structures which are primed with expectation (prediction) effectively look out into the noise of their immediate surroundings so as to identify or ‘select’ pertinent patterns from the noise. However, now it can be specified that at any point in the hierarchy, what counts as ‘immediate surroundings’ is the informational body of the level below it, which constitutes the informational environment in which a particular specialist structure subsists. In other words, what counts as pattern at any particular level of processing is noise from the perspective of the next level up.

There are two corollaries to this. The first is that noise (qua the global body of pattern) permeates the entire system. And second is that the entire system, level upon level, is continually working as it ought to (by design) via a process of selecting from noise (i.e. the body of all patterns) only those patterns which it needs, expects or wants to find, and in so doing automatically rejecting or deselecting all other possibilities, as the system at every level strives to negotiate its world.

It might be asked, what exactly is a specialist structure? And the answer is that it depends on which bit of the system we’re talking about or the perspective one takes. For example, if we’re talking about a human eye, then specialist structures are in the light sensitive cells behind the retina; if we’re talking about a human tongue, it’s the small clusters of cells known as taste buds; if we’re talking about the autonomic nervous system, then specialist structures are dispersed in nerve fibres throughout the body; and if we’re talking about a person, then it’s the bit of his or her personality that’s learned to value and so spot criteria such as sunlight, a sundae, a clause, and/or a beer can hat. These are very different specialisms. Yet one thing which unites them is that they all have a job which they are motivated to perform. And again referring back to the previous outline of prediction error theory, it was explained that their job is to be continually skimming and scanning the body of information (noise) which comprises their particular surroundings so as to extract the pattern(s) they predict (need, expect, anticipate) they ought to find. And they execute this task by effectively asking one simple question: Is it important? of every bit of would-be pattern which comes by. Whereupon, in feedback from the noise, and in the syntax of prediction error, the structure receives one of three replies: No; or Yes, maybe; or Yes.

What do these answers mean? No, means that prediction error is too great. That is, there is not enough correlation between the prediction and the likely bit of content, i.e. not enough pattern; too much noise. Consequently, motivation is withdrawn or diverted away from that particular likelihood, whereupon the would-be bit of pattern is
ignored. What happens then to the would-be bit of pattern? It depends. Perhaps it continues to circulate in the noise; maybe to be reselected at a later date; possibly in different circumstances; perhaps after adapting; or maybe it just falls away, i.e. breaks apart and drops backwards to be incorporated as noise into the informational level(s) below.

Yes, maybe, means that prediction error is lowered, but still there is too much error in the choice. That is, there is some important correlation between the prediction and the likely bit of content, but the sample is still too noisy for the system to engage with it. This stresses the system. Why? Because this is a deeply motivated problem-solving system which needs to understand its surroundings in order to work, i.e. survive, and in this situation it has hit an obstacle. So to deal with the obstacle extra energy, i.e. motivation, is made available to further the processing of that particular likelihood. As a result of which, the would-be pattern is kept alive, impotent but buoyed up, in the body of information of that particular level, which also makes it available as noise to the next level up. What happens then to the would-be pattern? Again, it depends. Perhaps it will just remain circulating at the same level before the system loses interest and it unobtrusively falls away. Maybe as other contents change around it, it will (in its original or an adapted form) become more relevant and so gain purchase at that same level. I say more about this in a moment. Or perhaps as it circulates, in a body of information which is noise from the perspective of the next level up, in part or in whole it will be selected from above as a (likely) important pattern, and in so doing the information it contains will be pulled up into the higher level. And if the latter, so the sequence repeats again.

So in this way, across time and multiple rounds of testing, the consequence of these two question and answer processes is that patterns of information deemed to be important are progressively selected-in, while the other stuff, deemed to be noise, is progressively selected-out. In other words, noise, qua the global body of informational patterns, ‘is continually being ‘funnelled through the reducing valve of the … system’ (Huxley 1954 [2004] p. 11).

Finally, Yes, means that prediction error is minimised. That is, the correlation between prediction and the report feeding back is sufficiently narrow for the system to conclude that the two factors are (likely, sufficiently) the same. What happens next in part depends on the pattern at point. For instance, if the pattern informing the system concords with the immediate or imminent occurrence of a basic highly typical stressor
(danger, threat), then the Yes response will be low in the system, motivation is going to be great, so that the stimulus itself ought to be sufficient to prompt an immediate reply, e.g. FIGHT! FLEE! NOW! However, putting aside this contingency (though it is important, and I return to it later), on receiving a Yes response, the motivated problem solving system takes a moment to rest on its laurels. This affords sufficient energy to prevent the pattern from falling back into noise, but not enough energy to have it contend dynamically at that particular level, and so not enough energy to have it pushing for selection at the next level up. This levelling out of energy and movement might be thought of as the pattern gaining lateral purchase at a particular level of processing, with the upshot that the relative stability affords the pattern a (transient) position in the brain’s own story.

And, by way of comparison, a different way of thinking about the levelling of activity which I suggest happens with a Yes response might be in terms of (neural) attention (Cohen and Dennett 2011), where attention in this sense is not personal awareness (Lamme 2003), but a processing function whereby a pattern of data is effectively held steady in the machineries of the brain.

So thinking about this in more Dennettian terms, the dynamic forces of the No and Yes, maybe responses as outlined above, offer a finer-grained explanation of what occurs at the level of (basic) belief, i.e. far below the level of the personality, as the system works to sort competing patterns of information, before the Yes response provides the explanation of how it (the system) selects or opts for one (most likely) possibility, that is to say, ‘places a bet on the truth’ (Dennett 1978 [1981b], 1987 [1989b, 1989c], 1991, 1991 [1993]). Furthermore, in a complex living system which is actively engaged in testing possibilities across multiple stages via countless specialist structures, at any moment there is not going to be merely one Yes in play. Rather at different places, re different contents, and different levels of processing, there will be innumerable Nos, Yes maybes and Yeses all jostling for position, which is an image which nicely supports Dennett’s idea of ‘spatially and temporarily distributed content-fixations in the brain’ which are under continual revision or ‘redrafting’ and which ‘yield, over the course of time something rather like a narrative stream ...’ (Dennett 1991 [1993] p. 113).

One final point about a Yes response is that the stability or ‘lateral purchase’ which it seems to afford is illusionary and always transient. That is to say, as circumstances across the hierarchy are continually changing, so the flow of energy and information up and down through all of its stages is forever on the move. Some
contents change (very) rapidly – a (very) fast turnover being typical in the roots of biology. Some contents change (very) slowly – a potential for longer slower changes being characteristic of certain higher-order competences. Yet always nothing is certain. It is a game of likelihoods. And nothing is permanent. As Dennett says, the brain is continually redrafting its narrative. And while Dennett talks about this latter point most explicitly in connection with experience (consciousness) (e.g. Dennett 1991 [1993]; Dennett and Kinsbourne 1992), as observed in chapter one, the principle is not limited to consciousness per se (cf. Dennett 1981b; 1991 [1993] pp. 457-8). And certainly, with the inclusion of prediction error reporting into the Dennettian framework, it becomes evident that the dynamic behind Dennett’s multiple drafts model of consciousness is continually at work throughout the system at every point of function. Hence, the observation made in chapter one, that when it is combined with the principles of prediction error reporting (at least, as the merge of the two theories is tackled here), so Dennett’s multiple drafts model of consciousness, itself is redrafted into a multiple drafts model of mind.

How a D-PE system finally opts to move this way or that

If the explanation that I have offered so far holds good, then the situation is such that the principles of a Dennettian framework and the principles of prediction error reporting have been merged to create a single Dennettian plus Prediction Error (D-PE) account, right up to the point where the brain’s own narrative is under continual revision through the pressures imposed by endless modifications and motivations. However, the narrative is just that, a story – albeit a good informational story. So, how do we move the explanation from mere narrative and into actual behaviour? And the answer is precisely in accordance with the principles of prediction error minimisation and what Dennettian theory says concerning probes.

I introduced the idea of probes or prompts in chapter one. Essentially a prompt is a stimulus, such as a natural object or a direct question, the onset of which in the wider environment demands an appropriate response from the system. And in light of what has been said about the system’s dispositions and motivations, so one way of thinking about how this might work is in terms of (functional) stress.

To explain. If the arrival of a prompt in the environment alerts the system that this is a something which (likely) needs to be dealt with (now), then it seems reasonable to think that this could stress the system out of its baseline ‘housekeeping’ mode and into
an alerted ‘responding’ (as it ought to) mode; so freeing a pattern of energy and information stored in a pertinent part of its (brain’s) narrative into the wider system; so triggering a pattern of apposite changes which (ought to) eventually effect a pattern of downstream behaviours which the system is betting will make the tension within itself go away. Hence, the instant Dennett (his system) witnesses that the sloop he is being offered is purple, the onset of the stimulus stresses the system into releasing from its (brain’s) own story a pertinent pattern of (negatively) motivated changes which effects a complex of downstream changes to the end that Dennett exclaims: ‘No! Not purple!’ And phew, danger over, the system relaxes into ‘housekeeping’ mode again.

Yet given the way in which the brain’s narrative is always on the move, so content available at T1 may not be available at T2, or if available may not be so in the same way. Hence, Dennett’s comment that the same prompt at different times may elicit a different response (Dennett 1991 [1993] pp. 113, 135-6 etc). This is not to suggest that on a different day Dennett would grab at the option of buying Flight, the purple sloop. He might, but I doubt it. However, his response could be a more measured: ‘Hmm, a purple sloop is not quite what I had in mind.’ Likewise, consider what Ted Hughes says about his approach to recording the events which he describes in the poems of his book Moortown Diary. Basically, Hughes wants a way not of distancing himself from but of ‘staying close’ to the actual event:

This sort of thing had to be set down soon after the event. If I missed a moment – which meant letting a night’s sleep intervene before I took up a pen – I could always see quite clearly what had been lost. By the next day, the process of ‘memory’, the poetic process, had already started. Though the details were still absolutely fresh, most of them no longer seemed essential to the new pattern taking control. (Hughes 1979 [1989] pp. x-xi.)

In other words, with the passing of time the content is pulled upwards into higher increasingly cognitive levels of processing where informational patterns are shaped and work quite differently to those associated with the actual event. Basically, it’s the difference between putting a forkful of omelette into your mouth and going: Mmmm! Then saying, ‘This is good.’ Then later that evening reflecting on the omelette as well as other omelettes from your childhood. And then many years later writing a treatise on the omelette as a way of life (cf. Martin J.R. 2001 [2006]).
This is a point of relevance for the following discussion of delusion. For the reason that in any medical consultation clinicians rely on patients to recall and express in their own words the onset and development of pertinent, but often hard to describe, unfamiliar, and/or confusing events. Whereupon, not only does the patient’s background work as a selective control on factors such as what actually counts as pertinent, but as Shitij Kapur observes, ‘Given that most patients come to the attention of clinicians after the onset of psychosis, phenomenological accounts of the onset of psychosis are largely anecdotal or post hoc’ (Kapur 2003 p. 15). Moreover, the circumstances in which a patient recounts his or her story are often stressful, not least because it is being recounted to a clinician or clinicians. Furthermore, on hearing the patient’s recount, the clinician then needs to select and record what he or she takes to be pertinent patterns in case-notes, which at a later time they or others may return to, consider, discuss, decide on, and maybe review, and then at a later date still the case might be reframed as a medical case study, and/or represented as data in academic research. The consequence of all this being that information concerning a patient at T1 inevitably changes through time at T2, T3, T4 and beyond via (increasingly) intellectualised levels of processing into (increasingly) intellectualised (i.e. ‘unreal’) realms. And in chapter four, I return to this point especially concerning Patsy and Mr P.

Still it might be queried that, okay, this is all well and good, but Dennett and Hughes are highly intelligent agents who may well have the capacity to negotiate their circumstances and choose, but there are more basic issues at stake, for the reason that all of the above is said to depend on biology, its structures and functions, but: One, how can what is essentially a blind mechanical system opt for the contents of a treatise, or for that matter the verses of Moortown Diary or the content of Consciousness Explained? And two, if it is prompted in a situation involving two identical or two equally unfamiliar stimuli, such as happened in the case study re the sloops Unicorn and Molly B, then how does a purely functional mechanism select and so motivate the system re what it ought to do next?

So in answer to the first query which is how can a blind mechanical system opt for the contents of a treatise, or for that matter the verses of Moortown Diary or the content of Consciousness Explained? I would begin by pointing out that the process I am outlining entails not just one but innumerable rounds of prediction error testing of contents, which low in the system are deeply typical and utterly non-linguistic, i.e. deal with the rudimentary stuff of life, but then working upwards, while basic informational
contents are still fundamentally non-linguistic, they start becoming more and more *about* language. That is, their patterns start reflecting the higher-order concepts of natural language through which a subject’s mind works to capture the significant features of his or her world.

Moreover, it can be reasoned that, initially, these linguistically coloured contents are not going to be those of polished discourse, but will present as crude rudimentary elements, which by way of rounds of error minimisation gradually coalesce into more usable forms. This may be illustrated by using an excerpt from literature which, as I explain in a moment, is not without justification. It is an example of the semi-formed unbroken stream of ‘sleep-talk’ of James Joyce’s novel *Finnegans Wake*, and it captures the subject’s first impression on entering a cluttered room, a study.

My wud! The warped flooring of the lair and soundconducting walls thereof, to say nothing of the uprights and imposts, were persianly literatured with burst loveletters, telltale stories, stickyback snaps, doubtful eggshells, bouncers, flints, borers, puffers, amygdaliod almonds, rindless raisins, alphybettyformed verbage, vivlical viasses, ompiter dictas, visus umbique, ahems a aahs, ineffible tries at speech unasyllabled, you owe mes, eyoldhymns, fluefoul smut, fallen lucifers, vestas which has served, showered ornaments, borrowed brogues ... (Joyce 2012 p. 183.)

Undoubtedly, as far as the rudimentary stages of linguistically coloured processing in the system go, this content is still linear and eloquent, i.e. pretty high-up. Nevertheless, the excerpt serves to illustrate the idea of a streaming litter of noisy semi-formed (on the face of it, unfathomable) linguistically coloured contents which the D-PE model predicts ought to be going on in the background as any intelligent subject negotiates his or her world.

Nevertheless, at some point these contents may start being tested at levels which are potentially accessible at the level of the personality, that is to say coalescing into something closer to the spontaneous linear-ish but still noisy and choppy trains of content (experience and thought) which are familiar to a subject as he or she goes about his or her days, engaging with his or her world. So now imagine walking into an unfamiliar study, and at the bewildering sight of surfaces littered with food stuffs, and books, and drafts of manuscripts, and letters, and bills, and more bills, and soot, and
smokers paraphernalia, and cigarette butts on the floor, stopping short with the
exclamation: ‘My word!’ And with it, we have stepped out of *Finnegans Wake* and into
the meandering ‘variegated jumble of images, decisions, hunches, reminders and so
forth’ which constitute the first-person stream of consciousness of another of Joyce’s
novels, *Ulysses*. And it is in explicit recognition of the parallel to James Joyce’s novels,
and specifically to *Ulysses*, that Dennett dubs the architecture of the living human brain

Furthermore, once content is available at the level of the personality for
everyday consumption, it might be ignored and so lost, or redrafted, or used then
rapidly dropped, or perhaps be recognised retained and polished. Whereupon, if one is
Ted Hughes or James Joyce or Daniel Dennett, then the effects of this polishing may in
due course realise a work of literature or philosophy while for the rest of us it may be
made specific enough to secure a fancy lunch.

I want a two-egg mushroom omelette, some French bread and butter, and a half
bottle of lightly chilled white Burgundy. (Dennett 1987 [1989] p. 20.)

Yet always there is the risk that we fail to hit the mark, and perhaps either underspecify
or be prompted to overspecify more than we really want to or should.

“I’d like some baked beans, please.”

“Yes sir. How many?” (ibid.)

And again we can observe the familiar pattern as contents move upwards into higher-
order levels of processing, and the inevitable progression through time into (over)
intellectualisation and the unreal.

So now turning to the second of the above queries which is that in a situation
involving two identical or indeed two equally unfamiliar options, how does a purely
functional mechanism select what it ought to do next? In response, I suggest three
possible courses it might take.

The first possibility is that the machinery ‘seizes’, i.e. arrests, interrupts or slows
its output, as the system works out across itself what it ought to do next. From the
perspective of a normal healthy human subject, this might be experienced as a moment
of ‘brain lock’ (blankness) or puzzlement (confusion) which leaves one grasping at
nothing for a moment or two, before one’s thoughts regroup and everything continues as it ought.

Second is that the brain simply selects one option, it has to, because this is the way in which a living system works. That is, if need is great enough, and so motivation sufficient, then given the circumstances the body will do something to meet its requirements, period. Moreover, should tension be driving the system hard enough, then the system’s jump to that something might be interpreted as impulsive, or disruptive, or perplexing by commonplace standards. Yet understandably it is going to happen. Then perhaps, just perhaps, there might be a subsequent round of reorganisation, rationalising, and maybe apologising for what just occurred.

Third is simply the practical point that it is highly unlikely if not impossible that two criteria are identical. There are going to be differences. For example, re the sloops *Unicorn* and *Molly B*, the most obvious dissimilarities are that the names are different, that by necessity they are moored in different locations, and so by necessity Dennett is going to observe them differently at different times. This is a non trivial point. For the reason that across multiple levels of processing in a system which is repeatedly striving to identify perhaps quite small evidences so as to meet its expectations, so small dissimilarities can rapidly add up.

Yet this itself might prompt a further query that, okay, but if so then why doesn’t the whole setup go wrong? That is to say, if a brain is disposed to leap, and if small differences can have substantial effects, then across an individual, let alone a population, would it not be the case that all manner of conclusions and behaviours are going to take hold? The answer to this query is yes and no.

Yes, because we can be confident (and not to say thankful) that in the warp and weft of even the most normal system across time offbeat conclusions are going to happen, for the reason that this is an informational model based on extracting (mere) likelihoods from (all-pervasive) noise. That is to say, uncertainty or unpredictability is built in. So why should we be thankful? Because, (a) offbeat conclusions are the engine of evolution, and so without them goodness knows where any of us would be; and (b) I submit that life would be pretty monotonous without idiosyncrasies getting periodically thrown into its mix.

However, no, and for two reasons. One, this is a deeply normative system which from its roots upwards is engineered to do the right thing. That is to say, ‘nature has built us to do things right; look for systems to believe the truth and love the good’
And two, it is a wide system. Accordingly, and in principle at least, all of its moving parts (including, for instance, other individuals, family groups, or a wider population) are implicitly working together as a whole so as to keep the entire organisation going as it ought, and in so doing each part effectively motivates, moderates or acts as a constraint on the rest. As Dennett says, on some level one gets by (more successfully) by way of ‘navigating the shoals of interpersonal relations’ (Dennett 1991 p. 50).

Yet, this said, it is more than evident that in particular cases things can, do, and indeed eventually must go wrong. That is to say, some part of a system, for some reason, may be pushed into working beyond its norms, i.e. outside of its tolerances, and with consequences that follow. For instance, certain drugs can addle matters throughout the system; organs can fail; and even in sober heathy people minds can become very, very confused. I say more about this when I talk about delusion, in chapter four.

**Competing patterns in noise**

As was observed previously, representing a wide deep Dennettian system via a simple hierarchy of six stages is on the face of it a massive oversimplification. It is an abstraction, a model, a story, but I submit that as far as stories go it’s a good one. Why? Because it is not a bit of mere whimsy, but instead it looks out into the environment from a particular perspective and for a particular reason and in so doing picks out a pattern which really does seem to tell us something of meaning about certain significant relationships in the world. To extend Dennett’s metaphor, it is a story designed to navigate the shoals of possibilities which was derived by navigating the shoals of possibilities. I suppose this makes it not entirely unrelated to a traveller’s tale: *Here there be monsters.* Likewise, consider Friston et al. when they write:

> Our capacity to construct conceptual and mathematical models is central to scientific explanations of the world around us. Neuroscience is unique because it entails models of this model making procedure itself. There is something quite remarkable about the fact that our inferences about the world, both perceptual and scientific, can be applied to the very process of making those inferences ...

*(Friston et al. 2006 p. 70,)*
Importantly though the model lays no claims to truth, for the fundamental reason that *ex hypothesi* there is no fact of matter for it to lay claim to. Consider that a neuroscientist, a biologist, a philosopher, a poet, a novelist, or Sally or Jim, could each potentially look directly onto the same bit of noise and each one select a different pattern (or maybe the same pattern at a different depth) as a meaningful representation of what seems to be going on, but with nothing definite to decide between their ‘rival’ descriptions of pattern in noise – not even time or an ostensible direct hit, i.e. predictive success. For the reason that small local victories are just that, i.e. small local victories, which get subsumed into noise when one steps back and takes a different (narrower, wider, alternative) viewpoint, and then steps back again, and again and so on. Whereupon, not only might the ostensible victory no longer be a part of a different (bigger, deeper) ‘victory pattern’, but it may no longer be recognisable as (a related) pattern in noise at all (Dennett 1991 pp 44-9).

[T]here could be two different, but equally real, patterns discernible in the noisy world. The rival theorists would not even agree on which parts of the world were pattern and which were noise, and yet nothing deeper would settle the issue. (Dennett 1991 p. 49.)

To put it a bit more sentimentally, we are each and all trying to make sense of the noise as best as we are able, because we are motivated (need, want) to do so, and as a result of which some of our patterns seem to work well (enough), whereas without doubt others seem to be pretty ... ‘dodgy’ ... in which case at some point the dodgy patterns ought to be challenged and so selected-out. All the same, in certain circumstances, this might not happen, or it might not happen as it ought. Whereupon, predictably, related patterns of dodgy downstream behaviour of a pertinent sort may arise. And, perhaps unsurprisingly, this is a point I return to re the cases of delusion in chapter four.

Even so, the point might be pressed that I have acknowledged that a six-stage hierarchy is a huge simplification, so one might query what happens if more complexity is factored in? And in reply I would say that actually, quietly, more complexity has been factored in.

For example, the way in which the environment has been defined comprises the actual physical surroundings, then the interpersonal, social and cultural surroundings, all of which house a pervasive body of informational pattern in noise. This to me seems
pretty complex. Then there is the human body, which is a staggeringly complex bit of biological machinery. Then motivation is not just motivation, but has been talked about as having different levels of urgency in re different needs and different patterns in different circumstances. However, to underline the idea that complexity is already inside the model, think for a moment about the stage of belief. This is the level at which data is moving through the entire system in ever-changing informational patterns and is doing so via a mind bogglingly wide deep and complex network of pathways and connections. For instance, I couldn’t even speculate as to how many connections there are in a living human brain alone. Yet within this framework every one of those connections is held to be on the same continuum as the one we’ve been talking about, in that each one will have stages below it (supplying noise), and stages above it (available for further testing if required), as it attends to the patterns of information it deems (expects, predicts) to be important, while ignoring or selecting-out the rest. And all the while, the brain is working, adjudicating these changes, as it drafts and redrafts the ever-changing contents of its story about the world.

And so a final query might be, isn’t all of this horribly solipsistic? And the answer is that it would be, if it were an account about a brain sitting alone in the darkness at the top of the control tower, as opposed to its having a parallel role in a wide connected information sorting system in a world where Dennett likes boats and Sally loves Jim.

Thus with a combined D-PE framework in position, I shift perspective from the underlying mechanics of a combined Dennettian plus prediction error system, to thinking about things from the more fluid standpoint of the subject, Dennett. First, in regard to Dennett making up then (perhaps) changing his mind to buy the sloop Molly B. That is, as he stands in a boatyard, how and why Dennett arrives at a specific conclusion re the circumstances of his world, and which (conclusion) he may subsequently revise. And second, thinking more specifically about experience, what happens when first Dennett sees the Molly B. Then, putting all of this together, I carry forwards the D-PE model to see what (if anything) of value it can say about how and why things develop as they do for each of the subjects in the cases of delusion in chapter four.
How Dennett Makes Up then (Perhaps) Changes His Mind

With the structure of the D-PE model in place, I turn to the normal case study of Dennett and of the Molly B, so as to scaffold an explanation from the perspective of Dennettian theory combined with the information sorting dynamics of prediction error theory as regards how and why a system, qua subject, namely Dennett, wants, and plans, and eventually decides to buy a boat, before (maybe) changing his mind and so electing to renege on the deal.

The rationale for the approach is that the explanation re the normal case study serves as a control, effectively telling us what ought to happen when a subject makes up and then changes his or her mind, i.e. when everything concerning the system works well (or rather, well-enough), so that if the explanation is then carried forwards and incorporated into a parallel case where things fail to turn out quite so satisfactorily, that is to say, into the abnormal cases of delusion in chapter four, then it provides a way of pinpointing how and why in the anomalous cases things could (plausibly) be going wrong.

Dennett’s vague unspecified hankering

This was explained earlier in the chapter when I spoke about the living system, and its reasons and motivations, but here in the case study I run through the main points again. We can start by observing that at base the subject Dennett has a lot of biology, which in order to continue as biology demands that certain basic needs are continuously met. For instance, there is a need for sustenance, and a need for personal contact, and a need for freedom of movement so as to acquire or avoid the boons and harms of its world. Low in the system, in the roots of biology, these needs are met by highly typical criteria via fast spontaneous or ‘brute’ changes to biology itself. However, the subject Dennett is more than biology. In addition, his system (he) has higher-order levels of intelligence, that is a capacity to retain, evaluate, abstract and even fabricate informational patterns re criteria which are (likely) desirable or detrimental to the integrity of its (his) being. Now factor in that Dennett is situated in an environment in which boat-related concepts are not only available, but are judged to be meaningful, and, specifically, to have positive value. So in consequence, and fundamentally outside of the remit of personal awareness, the mechanisms of Dennett’s own inferential machineries are likely to be selecting-in as important certain boat-related patterns – and so selecting-out competing patterns,
because a system can’t have it both ways – while pulling the valued contents up towards the higher cognitive levels of processing, because that is the level at which they have meaning. And all of which uses energy. So that if prompted what eventually gets released is not an explosive brute reaction, but a weaker more indeterminate agitation or hankering in the direction of boats, and about which at some point the subject, Dennett himself, may become aware.

How Dennett’s list of properties is derived

As observed above, Dennett’s system is motivated to accrue information regarding the patterns which it needs/expects to serve it well in some way. In large part, his system adapts itself to do this via its spontaneous motivational mechanisms, either by changing itself (e.g. by redrafting its needs and expectations), and/or changing its relationship to the world (e.g. by adapting its position in the world) in a relevant way.

Accordingly, if we shift our explanatory focus and think about this from the perspective of the subject, that is to say, the person, Dennett, we can expect that Dennett will (at least to start with) execute certain spontaneous intentional behaviours about boats. For instance, one day, uncharacteristically, he might pick up and peruse a boating magazine without ever thinking as to why. Yet, given the remit of the case study, we know that sooner or later, Dennett, at the level of personal awareness, gets in on the act. Whereupon, Dennett starts to notice and think about boats. And this is reflected in the case study when Dennett works at ‘narrowing his options’, i.e. working out his preferred list of desirable properties concerning boats.

So how does this happen in a D-PE model? Imagine that early on in the story the Dennett system has acquired access to a pretty eclectic set of basic informational contents re boats. Moreover, it is a belief set which is continually being added to and amended as it progressively adapts in relation to the ongoing circumstances which Dennett encounters as he negotiates his world. Overall this makes the belief set a big unwieldy shifting body of boat-related noise. Yet remember, noise is not mere noise, but noise is the body of patterns, and so noise is the medium in which patterns survive and move around. Hence, within the boat-related noise, the patterns a and b and c are circulating, but so are very many other patterns, such as patterns *a *b ab ac *c aba bby and so on in an ever-changing melee of basic belief. And all of this in a system which (by design) has the unrelenting job of having to decide.
Thus (metaphorically speaking) the system draws a deep breath, takes what it already knows is important (either by its hardwired mechanisms, or softwired learning) and uses this (information) to predict what it ought to find ‘out there’ in the noise. And not just once, but countless times over. And with each spurt of prediction, the system receives feedback in the syntax of prediction error reporting, and on account of which (feedback) it adapts its responses, maybe dropping that bit of pattern, or attending to it, or making it available for further testing, or by adapting its expectations in light of the ongoing flow. And in so doing, the entire system is continually working to maximise its predictions by reducing extraneous (to its needs, expectations) noise, which is to say error.

It has already been observed that deep in a system’s biological roots, where predictions are hardwired, there is naturally very little room for manoeuvre re the system’s responses, but working upwards to where there is more plasticity re predictions and responses, there is correspondingly a greater capacity for choice. Accordingly, we can now turn to Dennett, sitting at his desk with a blank sheet of paper in front of him and a pen in his hand as he issues to himself the following prompt – which effectively and spontaneously creams off a topmost pattern of content from the brain’s own narrative in response.


Importantly, Dennett is not merely a passive recipient of his own impersonal processing, since once he is aware of what his options are (i.e. the possibilities available), he (almost certainly) takes an active role in minimising the noise. That is, at the level of the personality Dennett draws on what he knows to test his own hypotheses, expectations, prejudices, preferences etc. against the noisy but evolving list of properties in front of him, and in so doing he assumes a personal (or first-person) role in the predictive hierarchy – selecting-out undesirable patterns and selecting-in patterns which he deems have positive value – until finally he, Dennett, settles on a particular pattern of content: a sloop with the properties a-b-c-q-r-s.

Clearly the list is not comprehensive. And could it have turned out differently? Yes, for sure. In different circumstances, which includes at different times, different contents will be fixed in the (Dennett’s) brain’s own story. And while some patterns
may be strong, and so likely to be selected-in in almost any normal circumstance. Others will be weaker, and so occasion less certainty as to whether or not they make it into a draft. And yet others will be so tangential that they will have been selected out by the system (very) long before Dennett’s list was ever a feature. We might think of these latter sorts of pattern as patterns *in potentia* held deep within a brain’s ‘virtual library’ (Dennett 1991 [1993] pp. 359-362). Perhaps such information might relate to the would-be pattern *purple-and-sloop*. (In the game of likelihoods, who or what would deem a purple sloop to be a viable possibility?) Nevertheless, when presented with the stimulus of the purple sloop *Flight*, deep within Dennett’s system alarm bells sound, the pattern is conceived, and given some lightening fast rounds of highly motivated redrafting, the property of being not-plain is swiftly and explicitly appended to Dennett’s list.

*At the boatyard, Dennett interacts with Bob – and the sloops*

Moved by motivation and direction, Dennett arrives at the boatyard, where he is met by salesman Bob. Bob is an intelligent human agent with his own motivations, gen and strategies. So that when Dennett explains his requirements, Bob is moved to respond with some typical salesman blather: ‘That’s great! I have three boats which meet all of your requirements. They are *Flight*, *Unicorn* and the *Molly B*.’

*Flight* turns out to be purple, and we know what happens next. Then Bob shows Dennett *Unicorn* and *Molly B*, the two tantamount identical sloops, but there are (necessary) differences between them. For instance, Dennett witnesses the boats at (marginally) different times, and/or in (marginally) different relations to each other and the surrounding environment(s) which he and they occupy, plus quite obviously the sloops have dissimilar names. And in a massively connected system which is driven to choose on the basis of its expectations even small differences can have (or amount to) significant effects.

For instance, maybe without Dennett ever drawing the explicit connection, his background fondness for *Ulysses* and/or (perhaps) *The Beach Boys* could positively colour the importance and value awarded to certain patterns of information deep inside the predictive mechanisms of his system. Whereupon, the conclusions which it (the system) draws in various places across itself will be coloured or biased in a *Ulysses* and/or *Beach Boys*-ish direction. Moreover, as it is precisely these conclusions which supply the noisy data for the next round of processing, so there is an increased likelihood that in the next round of testing the contents of these patterns will (in whole or in part) again be
selected in – upon which by necessity competing patterns will again be selected out. And thus it goes on. So that via ongoing rounds of processing, patterns deemed of importance and value (by the system) get progressively maximised and patterns deemed extraneous (by the system) get progressively minimised. Accordingly as to Dennett’s system’s original *Ulysses* and/or *The Beach Boys* inspired biases, we can say that if motivation is maintained long enough, and/or the relevant or related patterns are valued enough or prevalent enough, then via a massively disseminated but coherent succession of connected changes, contents related to those patterns are more likely to acquire (transient) purchase in the brain’s own story. Whereupon, if other conditions hold, Dennett may quite reasonably observe that he prefers the *Molly B*.

All of this plays out deep inside the (Dennett’s) system’s problem solving machineries. Nevertheless, as anyone who is disposed to, say, blushing probably knows, deep motivational responses can have some observable effects. Without doubt, Dennett knows this too. Thus he almost certainly tries his best to remain as dispassionate as possible. Nevertheless, Bob is an experienced and motivated salesman, and he (Bob, or his system) has learned to discriminate even subtle indications of a positive response. Also Bob knows that he needs to act while Dennett’s motivation is hot. So he prompts:

BOB: So which do you prefer, *Unicorn* or the *Molly B*?

DENNETT: [Brain anxiously working so to meet social obligations is/feels compelled to respond] Err ... the *Molly B*.

All the same, Bob knows that there is a world of difference between Dennett preferring the *Molly B* and Dennett buying the *Molly B*. So again he applies a couple of honed salesman strategies to apply stress to Dennett’s system. One is the blatant falsehood, ‘It’s one of a kind.’ (And yes, it’s a lie, but even so the pattern of information is now explicitly circulating in Dennett’s environment.) Two is a goading prompt re limited availability and $$$. (In other words, it’s a sly provocation in the direction of: Act now! You have reasons for doing so! Otherwise you will miss out on this boon!)

*So what happens next? Dennett makes up his mind.*

So here we have Dennett, standing on the jetty, his system under stress, faced with one of those 50:50 choices. On the one hand, there is the desire to keep hold of the $$$. In
his pocket. On the other hand, there is the desire for freedom of movement and the status of owning the *Molly B*.

The outcome of this deadlock is perhaps that Dennett experiences what I previously referred to as momentary brain lock. That is, his brain temporarily interrupts its normal flow of processing to focus on working out this problem re its world, and in so doing committing to what it needs to do next. From Dennett’s point of view, this could feel like a short spate of anxiety or confusion as perhaps half baked options bat too and fro. And if the outcome of this spate of intense problem solving is that the brain derives a favoured conclusion then great, that’s the way to go, and the system moves on.

However, if there really is no clear way forward, then Dennett’s system will eventually resort to its alternative strategy and simply opt one way or another. It has to, because basic need and social mores demand that Dennett can’t stand rooted on a jetty forever. Moreover, it is a reasonable enough move by the brain, since if these really are the only options it can find, and if there really is no way it can optimise between them, so matter-of-factly one jump forwards must be as viable as the other. Accordingly, the brain engages in a final round of prediction error minimisation then driven by stress and prompted by need and obligation it jumps this way or that.

So let us say that Dennett’s system jumps in favour of buying the *Molly B*. This collapses the relevant pattern of possibilities held in the brain’s narrative into a related pattern of downstream changes which (may) in due course coalesce as a (typical) pattern of (potentially) observable linguistic behaviours. At which point in time, and at no time before, Dennett (may) explicitly entertain the thought: *Okay, I want it*, and (may) follow this by stating out loud: ‘It’s a deal, I’ll buy the *Molly B*.’ Why only ‘may’? Because even at this stage informational contents are subject to redrafting, and so it is possible they could adapt before or as Dennett holds (corrals, retains) them in awareness. This can be compared to what Dennett observes re the Orwellian and Stalinesque redrafting of experience (which I outlined in chapter one, and say more about re Dennett’s experience of first seeing the *Molly B* in a moment) for the reason that in the same system experience and cognition are behaviours on a single continuum and so in principle the same rules (should plausibly, unless a special story is told otherwise) hold.

However, let us say that Dennett explicitly holds onto the conclusion that, *Okay, I want it*. It is an act of personal assent re the (likely) veracity of that particular conclusion or ‘sentence’ (Dennett 1978 [1981b], 1991). It marks a hallelujah moment of
personal insight, since, as Dennett says: ‘[i]t is the occasion of the discovery I make about myself’ (Dennett 1978 [1981b] p. 303, see also Dennett 1991 [1993] pp. 193-9). And properly speaking it marks the moment where Dennett can correctly say to himself or the world: I have made up my mind (Dennett 1978 [1981b]) – as opposed to all the exertions and deliberations which have gone on previously but which have been fundamentally hidden to Dennett’s personal (or first-person) point-of-view (ibid.).

According to Dennett, and again properly speaking, the outcome, Okay, I want it, is not a belief, although it has its roots in belief (Dennett 1978 [1981b], 1991), but it is actually a species of judgement (ibid.), i.e. of ‘content fixation’, but not one like we have been talking about thus far which is buried deep and invisible inside the brain’s own narrative, but one which is, if you like, an outside and observable part of Dennett’s own narrative concerning what is (likely) going on in his (internal or external) world.

And certainly it is not the first time that Dennett has done this. For instance, it happened when one by one Dennett added to and adapted the contents of his list of preferences. It happened when prompted by Flight Dennett responded with aversion. It happened when Dennett observed his preference for the Molly B, and when prompted by Bob he affirmed this preference. Plus a thousand or more times and which together make up the ‘variegated jumble of images, decisions, hunches, reminders and so forth’ which constitute the first-person stream of consciousness of Dennett’s day (Dennett 1991 [1993] p. 212).

Finally it might be queried, okay, but is it (that is to say, Dennett’s own story) true? And the answer is that no it is not, it is a fiction. A story. Which is at best a version or a bet on the truth (Dennett 1987 [1981b], 1991). Nevertheless, all things being equal (by design) it ought to be a good (enough) fiction (e.g. Dennett 1987 [1989b, 1989c]), which indeed may be true (cf. Dennett 1991 [1993] p. 364), and which works to tell Dennett (or a second or third person observer) something meaningful about his (Dennett’s) view of the world. (Or, if the subject is a novelist, such as Joyce, then something meaningful about the unreal world which he or she is explicitly concocting in his or her work.)

**Dennett changes his mind**

However, as Dennett makes his way home and the positive stimuli of Bob and the boatyard and potential boat ownership recede, so the prospect of the chosen outcome, and not forgetting the hastiness with which it was opted for, may well motivate some
bothered rounds of reorganisation, rationalising, and maybe stressing about what just occurred. As a result, Dennett’s train of thought anxiously chatters away: should I ... do I ... ought I ... and so on, with each fretful sentence effectively prompting the next. Correctly speaking, this onslaught of ‘second thoughts’ does not constitute a change of mind – though it may well reflect mounting internal tensions and so indicate that a change of mind is on the way (Dennett 1978 [1981b]). Then back at home Dennett reflects further on the day’s events. He considers spending $$$ on a boat. He does a spot of extra research into mooring fees, maintenance costs etc.. He talks to his family, listens to their enthusiasms and concerns. Finally, let’s say the gas bill arrives. In itself, this is a trivial event. However, it is quite within the sphere of rationality, that the unexpected arrival of this ‘you owe me’ functions as a prompt which is sufficient to collapse in a particular direction all the motivation and information which has been mounting. Whereupon, Boom! Dennett assents to the sentence that: I don’t like the Molly B enough to buy her. And upon this, he telephones Bob to say, quite correctly, ‘I’ve changed my mind and no longer want to go through with the deal.’

This concludes the D-PE account of how a subject, Dennett, makes up and then changes his mind as regards the purchase of the sloop Molly B. Central to the account is the idea of pattern and noise, where noise is not mere noise but the body of all pattern. Plus the account tracks the explanation through the Dennettian hierarchy of stages, namely, those of environment, biology, need, motivation, belief, judgement, prompt, and behavioural changes, and some of which (behaviours) feed back as patterns of change in the informational noise which comprises the observable environment, and so become available for selection as pattern at a subsequent round of processing (interaction). And it traces how the entire system at every level follows a limited set of rules in a process which functions to maximise the patterns which are important to it (predictions) and in so doing reduce or minimise information which is judged to be extraneous noise (error), so as to afford the subject (or indeed any specialist structure at any level of function) a meaningful and coherent first-person narrative about his or her world.

Nevertheless, nothing so far has been said about Dennett’s experiences. Yet thinking back to the discussion of the prevailing theories of delusion, the role of (abnormal) experience was a significant issue. Also thinking forward to the discussion of delusion from the D-PE perspective, experience is also a consideration, especially re the
(very) strong hallucinations which characterise Mr P’s case of perceptual delusional bicephaly. Accordingly, before finishing chapter three, I turn to the other side of Dennett’s theory of mind to say something about how the account works to explain Dennett’s everyday perceptual experiences, and specifically Dennett’s experience of first seeing the Molly B.

When First Dennett Sees the Molly B

Dennett’s theory of experience as it was outlined in chapter one is that, driven by need as it goes about its day, a living system is continually working to derive a meaningful story about what is going on in its world as regards the various criteria in relation to which it needs to act. It is a story which is composed of myriad instances of ever-changing content-fixations or ‘judgements’. And many of these judgements will be fast spontaneous and relate to brute function, while others may be far slower and involve highly intellectualised themes. However, at a stage somewhere between bottom-end function and top-end cognition, Dennett proposes that there is a special class of judgements to which he gives the name ‘presentiments’. Presentiments are instances of content-fixation in the brain’s own narrative, and the bit of the story which presentiments presume to tell relates to meaningful objects and properties in the system’s, qua subject’s, ‘observable’ world. (Dennett 1991 [1993] pp. 113, 343, 364-5, 457).

Again from chapter one, it was observed that like all of the brain’s judgements, presentiments have conceptual edges, are derived at multiple different locations within the brain, are propelled by different strengths of urgency, are at various stages of drafting, and are inaccessible to personal experience, since their role is to inform the brain – that is, bodily organ – via its functional machineries about what is (likely) going on there, and there, and there, re the living system’s relationship to its wider surroundings (Dennett 1991 [1993] pp. 113, 365). However, among the brain’s judgements what makes presentiments special is that at least some of the concepts which corral them are of a kind which, if the narrative should be prompted in a particular way, release downstream behaviours of a kind about which the subject could (perhaps) come to be personally aware, i.e. experience, reflect on, argue about etc. (e.g. Dennett 1991 [1993] pp. 362-8).
So within the D-PE framework as it has been developed in this chapter, what can be said about the normal case in which Dennett first sees the Molly B?

I suggest that the key point is the idea of how the noisy content of the observable environment is narrowed or reduced in accordance with the living system’s expectations, in a ongoing process, a particular stage of which entails that, at very, very many different junctures across a living brain, contents get corralled via very, very many scrappy and short-lived coming togethers of (in part) bottom-up noise (error) supplying the available data, and (in part) patterns of expectation (predictions) fixing top-down edges on the noisy informational clutter of the level(s) of processing below.

An outcome of this would be that, if when visiting the boatyard Dennett (his system) is quite reasonably expecting to witness land, sea, sky, a salesman and boats, and assuming that the environment of the boatyard contains pertinent tokens of these patterns, then bottom-end to top-end Dennett’s system is likely and effortlessly going to organise its processing in light of its expectations and the data at hand. Round after round, honing its selections of the important and valuable stuff which it expects to discover, and in so doing progressively selecting-out more and more of what in the circumstances it takes to be noise. To the conclusion that lo, Dennett with neither glitch nor perplexity stands on the jetty and sees a boat.

Alternatively, it is likely when Dennett steps into the boatyard he really, deeply is not expecting to witness Elvis. So as per the D-PE model, even if Elvis were there, then there ought to be a remoter (but not zero) chance that Elvis-related contents are going to be (repeatedly) selected in, which means a remoter (but not zero) chance that Dennett would see The King. Unless, that is, a sufficiently pertinent prompt involving, say, sequins or Suspicious Minds were to drag forwards certain Elvis-related patterns which would be circulating impotently in the messy noise of Dennett’s global belief set. This can be compared to what Dennett says about Betsy playing a game of Hunt the Thimble (Dennett 1991 [1993] pp. 333-8).

Yet at this point it might be queried, but what about all the unexpected sightings of Elvis everywhere, for instance in bread products and similar objects? In response I would observe that down in the roots of the system there are fast hardwired highly motivated mechanisms whose job is to identify patterns of information concerning fundamental criteria which are highly typical and typically highly important in the (subject’s, system’s) immediate world. It is not implausible that the discrimination of patterns relating to (the features of) faces, is one such deep bottom-end task. If so, then
as bottom-up highly motivated noisy face-data makes its way upwards through the stages of processing it will be repeatedly tested by a series of discriminations via top down concepts as the system works with urgency to make sense of the competing patterns in noise. And re the case of Elvis, this sequence might be illustrated as looking like this:

Is it a harmful face? No.
Is it favourable face? Yes, maybe.
Okay, it seems favourable.
Is it a face I know? Oh Wow! It's Elvis.

Yet always the contents of a narrative are vulnerable to the pressures of reworking. That is to say, at any stage, any pattern of content fixation is vulnerable to the pressures of redrafting, either through pre-modification (i.e. upwards pressures via changes in the earlier level(s) of processing; cf. Dennett’s concept of the ‘Stalinesque’ revision of experience) and/or post-modification (i.e. via changes at the same level of processing or downwards from higher stages of processing; cf. Dennett’s concept of the ‘Orwellian’ revision of experience) (Dennett 1991 [1993] pp. 115-26; Dennett and Kinsbourne 1992). Plus given the ongoing flux of content up, down and across the motivated system, so different prompts at the same time or the same prompt at a different time could have a markedly different effect. Hence, in the case of witnessing Elvis’s face in a random object, one might return the next day and quite reasonably be disappointed.

Oh, Elvis is no longer there. I only see a brown splodge where yesterday I saw his face.

I return to these ideas in the following chapter, especially with reference to Patsy.

Finally, presentiments have the potential to collapse the energy and information they contain into the wider system in a way which works to inform various downstream processes, and in so doing (perhaps) eventually leads to conceptual – specifically, in line with the concepts of natural language – changes which are accessible to the subject at the level of his or her personality (e.g. Dennett 1991 [1993], pp. 321-68), and the utility of which is to literally inform the system at the level of the personality that something is (likely) the case in its world. So that, if other factors hold, relevant patterns of content
get selected-in to the subjective first-person story (ibid.). And, it is at this point, that Dennett standing on the jetty sees *that* large white sailboat situated in the centre of his perceptual field.

Importantly, in line with Dennettian theory, when Dennett sees the sailboat it is not as a result of an actual conscious metamorphosis whereupon all of the elements which comprise the experience of the *Molly B* (e.g., white + boat + mast + sails + ropes + jetty + blue + sea etc.) ‘come together’ in consciousness to present the subject, Dennett, with an actual integrated coherent view of the world in virtue of which there is literally *something-it-is-like* for him to see a boat (Dennett 1991 [1993]; Dennett and Kinsbourne 1992; Cohen and Dennett 2011).

Instead, the effect of releasing related patterns of energy and information into the system realises conceptual changes which generate something like a shifting multi-level highly fragmented phenomenological text which instructs across the level of the personality that, for instance, there is a sailboat ahead (Dennett 1991 [1993]; 2002a, 2003, 2005 [2006b], 2007 etc). That is, there is no actual picture presented to one in the colours of consciousness, and so literally *nothing-it-is-like* for Dennett to see a sailboat. Instead, there is ever-changing content relaying streams of information which is sufficient to ascertain from Dennett’s point of view standing on the jetty that there is *something-it-seems-like* to see that boat.

Concerning the phenomenological text. From Dennett’s perspective, it is an autophenomenological text which is being created. That is, the system spontaneously informs at the level of the personality that *x* is up ahead, whereupon from Dennett’s perspective for all the world it seems to him that when he looks he sees a sailboat with a mast, white sails etc. in the middle of his perceptual field (e.g. Dennett 2007). Alternatively, from the standpoint of a second person observer, the text is heterophenomenological. That is, if Dennett were to be prompted, ‘What are you looking at? What do you see?’ And he reflects and via the mores of natural language responds, ‘A large white sailboat.’ This ought to be sufficient to infer that Dennett judges that he is seeing a boat (ibid.).

That is to say, if Dennett says that he sees a sloop, and if his responses re said sloop are internally coherent and cohesive; and if his wider public behaviours re said sloop are similarly coherent and cohesive; and if we have no reason to think that Dennett is lying or engaging in fraud or otherwise being deliberately unreasonable; then what grounds does anyone have to claim that Dennett is not seeing a sloop? The answer
is no grounds at all. And again, I return to this point specifically re Mr P and the second head.

Finally, again it might be queried, in any instance is what Dennett sees true? And again, the reply is that it is a fiction. It is all a fiction. It is an informative layer in an epistemic story, or a bet on the truth (e.g. Dennett 1991 [1993] pp. 362-8). Nevertheless, by design, it is a layer of story which ought to be a good (enough) fiction, i.e. one which in any normal circumstances will be reliable enough. Why? Because in a system designed to do the right thing, and to be continually adapting its behaviours so as to do the right thing, then in whatever circumstances it finds itself, as best as it can manage, it will strive to do the right thing, time after time after time.
CHAPTER 4

A DENNETTIAN PLUS PREDICTION ERROR ACCOUNT
OF DELUSION

Introduction

In this chapter, I take the model of explanation developed through the normal case study of how Dennett makes up his mind, and subsequently changes it, re the purchase of the sloop Molly B, and apply it to an explanation of how and why a subject could make up his or her mind oddly, and then resolutely fail to change it in a case of delusion.

The cases of delusion in the order which I address them are. Erotomanic delusion, via the case of Patsy. Capgras delusion, a psychological explanation, via the case of Mme M. Capgras delusion, involving an organic pathology, via the case of Mr A. And perceptual delusional bicephaly, via the case of Mr P.

Concerning these cases, the key explanatory dynamic I explore is that in certain circumstances a fundamentally normal system can become vulnerable to the effects of noise, and especially to a pernicious type of slippery noise, which in the absence of suitable constraint isn’t rejected by the system’s inferential mechanisms, but tends to be repeatedly reselected-in, so that in time problems of a predictable and recognisable kind can develop in its overt behaviours.

In closing, I refer back to the two research questions which were identified in chapter one and consider how the D-PE approach as it has been developed here concords with both. Namely, re the philosophical problem of delusion as it occurs in a clinical setting: the first question asks which theory of delusion most comprehensively addresses the fact of delusion and its associated features; whereas the second question, which follows on from the first, asks which, if either, of the standard empiricist or rationalist explanatory models of delusion works the best.

And to end, I suggest two further of areas of research within the clinical sphere where a D-PE approach might be applied and contribute.

Finally, I would like to thank Professor Paul Noordhof for feedback on drafts of this chapter, and the Mind and Reason research group at the University of York, especially Dr Louise Richardson and Dr Louise Moody, for their support and
observations on a presentation I gave on an early draft of the case study of erotomanic delusion.

Three Cases of Delusion

Erotomanic Delusion

In chapter one, I described in some detail the case of erotomanic delusion. In short, erotomanic delusion is a disorder which is characterised by a subject holding the (false) conviction that *s/he loves me* of a person, i.e. the object of the delusion, who is typically (but not necessarily) of a higher social standing than the subject, for instance, a doctor or a celebrity (APA 2013 p. 91; Fitzgerald and Seeman 2002). The nature of the supposed relationship can vary. For instance, some subjects maintain it is a sexual relationship, while for others it’s an idealised connection (Menzies *et al.* 1995; Lloyd-Goldstein 1998; Kennedy *et al.* 2002). Hallucination is not a typical feature, but should hallucination occur it presents as a later factor and typically involves weak or dreamy content (e.g. Kennedy *et al.* 2002). The delusion can present as either a monothematic delusional disorder, or as a particular feature of a complex disorder, such as schizophrenia (e.g. Menzies *et al.* 1995; Lloyd-Goldstein 1998; Kennedy *et al.* 2002). The duration of cases differs, with some episodes of erotomanic delusion lasting a short time, perhaps only a few weeks; while other cases persist, but change progressively from object to object; and in others the subject remains fixed on a single object for a very long time (e.g. Kennedy *et al.* 2002; Fitzgerald and Seeman 2002). As regards comorbidity, there is no known significant correlation between instances of erotomanic delusion and either strong or weak neuropathology. Indeed, it would be characteristic of primary monothematic erotomanic delusion that a subject who presents with the disorder is otherwise normal, i.e. a healthy enough individual. On the other hand, antecedent social factors, typically involving social isolation, disadvantage or disaffection, are recognised to be co-occurring factors (e.g. Menzies *et al.* 1995; Lloyd-Goldstein 1998; Meyers 1998; Fitzgerald and Seeman 2002; Kennedy *et al.* 2002).

Thus the task is to provide a D-PE account as to why a fundamentally normal subject presents in a clinical setting with a complex of features symptomatic of erotomanic delusion, and moreover retains the delusion despite its implausibility. The explanation I offer is that an essentially normal need-driven subject (system) has to
negotiate a key environmental deficit. However, in doing so it (the system) becomes susceptible to noise. Yet a susceptibility to noise can, in certain circumstances, lead to a type of deep-seated ‘confusion’. And a consequence of this confusion is that certain predictable patterns of ‘anomalous’ behaviour can become apparent. Whereupon, a diagnosis of erotomanic delusion could eventually be made.

I present the discussion of erotomanic delusion through another fictional case study, namely, the abnormal case of Patsy and Handsome. And as with the normal case of Dennett and the Molly B, I track the explanation of the abnormal case through a series of stages. The stages are: Subjective need and environmental deficit – identifies the subject as a normal (healthy) human being, but who is caught up in a particular complex of special or difficult circumstances re a significant environmental lack; Prediction error and risk – suggests that the way in which the subject’s fundamentally normal problem-solving system handles the deficit is by opening itself to epistemic risk; (Slippery) Noise – explores how a system opened to risk is vulnerable to noise; A developing psychosis – considers how, given an absence of suitable constraint, (slippery) noise can agitate and confuse a fundamentally normal system in a very significant way; Cognitive (psychotic) insight – discusses how and why for the subject a delusion-type thought can coalesce; Hallucination – considers how and why the sorts of hallucinations which are sometimes reported in cases of erotomanic delusion could occur; Diagnosis – takes the D-PE account of the clinical case of erotomanic delusion to its conclusion, i.e. a clinical diagnosis.

Subjective need and environmental deficit

In chapters one and three I presented in some detail what it means to be a Dennettian subject (system), and especially so in accordance with prediction error theory’s requirement of the system as a connected hierarchy of functional stages.

In summary. A subject is approached as a living system which is organised by way of a hierarchy of connected stages, all the parts of which work individually and together to share energy and information as it (part, system, subject) negotiates the ups and downs of its environment with the fundamental objective of maintaining its individual (and so by extension the whole system’s) integrity through meeting its own (and so by extension the system’s) basic needs. Across the system, basic needs can be tokened in related but different ways. For instance, at the bottom-end of the system in the roots of its biology, the basic needs of all living human subjects are met by a limited
set of highly typical criteria, such as oxygen, nutrition, movement and personal contact, which the system moves very fast via spontaneous changes and behaviours to grab or avoid. Thus, at this level, there is no or minimal room for manoeuvre between the subject and her world. In contrast, moving up in the direction of the system’s top-end cognitive capacities, there is an increasing potential for the same basic needs to be satisfied across time via ever-weakening drives and via potentially novel criteria, such as a desire for a fancy lunch or a boat. So at this end of things there is progressively more room for manoeuvre, and with it an emerging capacity for personal choice, as the subject negotiates her world (e.g. Dennett 1984 [2002b], 2004, 2017).

However, an important point about a Dennettian system, is that it doesn’t just stop at a subject’s skin. Instead, the system pushes outwards to include the environmental surroundings within which the subject (and her parts) must act and interact. The idea is that every level of processing effectively takes place in its own environment. So from the perspective of the brain, the body is an environment. From the perspective of the body, or subject, the environment comprises her physical surroundings, interpersonal surroundings, and the wider social and cultural surroundings in which she must, by necessity, function. And a defining characteristic of any environment is that it is filled to overflowing with informational noise, where noise in this sense isn’t mere noise, but noise as the body of all informational patterns, and so the medium in which patterns survive and move around (e.g. Dennett 1991/1998 [1998b], 1991).

Accordingly, the system at every single level of function has a further basic requirement. It needs to minimise or reduce noise, i.e. the body of the information available to it, so as to select-in those patterns which are important to it and in so doing select-out the rest. And it is in this matter that I appeal to the principles of prediction error theory to tell a story about how the reduction of noise and so the selection of pattern could play out (e.g. Friston et al. 2006, Friston 2009).

One more point about environments is that they can be more or less ideal (e.g. Dennett 1984 [2002b] pp. 51-73, 1987 [1989c] pp. 49-51). The ideal case is one in which there is a perfect correspondence between what a system across all pertinent levels requires and what the world across all pertinent environments has to offer, whereupon the maintaining of the system’s needs should be a seamless state of affairs. Be that as it may, environments are seldom (if ever) perfect. Yet even in an average case an environment will tend to sustain well enough. For instance, all things being equal, a
subject doesn’t labour to, say, breathe, breathing just happens. Likewise, when Dennett almost bought a boat, the only real point of tension was at the prospect of handing $$$ to Bob. Not all subjects are even this well located, however. For instance, in the case of erotomanic delusion, the data suggests that here is a subject (system) that is likely to be having difficulty maintaining requirements which are decidedly more basic than a yen for a boat. Specifically, if, on the one hand, we have a system which has a basic need for personal contact. But if, on the other hand, here is a subject who is, so the data tells us, ‘shy, lonely, (sexually) inhibited, withdrawn, isolated; with poor levels of social, emotional and occupational support and functioning; and of unattractive physical appearance’ (e.g. Menzies et al. 1995; Lloyd-Goldstein 1998; Meyers 1998; Fitzgerald and Seeman 2002; Kennedy et al. 2002). Then it does not seem unreasonable to conclude that to some extent the latter condition is going to thwart the drive to satisfy the former condition. Hence, this is a system (subject) facing a significant lack.

But just what is lacking? This really depends on what one takes personal contact to mean. Consider that the pattern personal contact takes will depend on, first, the deeply typical biological machineries which motivate the system from the bottom-up, and, second, on the top-down conceptual edges which get drawn around the up-swelling tide of biological affect in a particular case. The outcome of this will vary somewhat between cultures and persons. This is because the thoughts a subject thinks, the conceptual edges she draws, are dependent on the beliefs that she (or her system) entertains, and which (beliefs) are influenced by manifold factors, such as biases, biography and background gen, as well as the pervasive atmosphere of prevailing (and possibly conflicting) cultural and social norms, i.e. ‘memes’, in which she has to survive (e.g. Dennett 1991 [1993] pp. 200-10 etc., 2017 pp. 205-47). So, for instance, across a group of people we could see ‘personal contact’ variously tokened as possession, sex, romance, enduring affection, or simply as friendship. Yet what this means is that a subject, or more to the point her system, is going to have meaningful expectations concerning the type of relationship it deems it ought to (be able to) find in its surrounding environment, and so the informational patterns relating to which will be precisely the patterns it is driven to hunt for in the surrounding noise.

As observed above, this is all well and good if needs, belief set and world coincide. However, taking the case of a typical subject who presents with erotomanic delusion. This is a fundamentally normal individual, with (we can assume) all the needs and drives of any other normal individual, but who is chronically caught up in the sort
of special and difficult circumstances outlined above, and who, in addition, (necessarily) lives in a culture of concepts concerning, for instance, glamour, and romance, and happily-ever-afters, i.e. the sort of stuff peddled in mass-market movies and celebrity magazines. Then, in this sort of case, the discrepancy between need and fulfilment is likely to be deep and messy and continuing. So that, if alternative options fail to be in place, then as the deeply normative system repeatedly misses its objective, so the unremitting deficit is likely to generate additional motivational pressures or tensions, which without relief will mount unremittingly, so that in time the system will be pushed into working under stress, i.e. outside of its tolerances (norms).

One might ask, does the deficit have to be chronic? And this is an empirical question to which I don’t know the answer. Nevertheless, it seems plausible that in certain circumstances, an acute stressor could serve as an unexpected prompt which could have a similar effect. A viable suggestion might be cases of bereavement and/or shock. (See case study 2b, re Mme M.)

However, note that, at this stage of the explanation, it is not the subject per se, but the way in which the system works or holds together which is said to be stressed. Indeed, it is reasonable to assume that in the early stages of erotomanic disorder, the subject will be uninformed that anything of significance is amiss. Yet the implication of this is that, even if it can be reasonably said that the root of the problem is located in the environment, i.e. in special and difficult circumstances with which an essentially normal subject contends, there is still an explanatory need to shift the account from a system under stress to a subject presenting with a delusion that s/he loves me. So to make this explanatory move, we need next to consider the effects of deep stress on the predictive processing capacities of an essentially normal living system.

**Prediction error and risk**

As previously described in chapters one and three, prediction error reporting is the process whereby a system (qua living body, brain) takes what it already knows, i.e. the hardwired expectations and the softwired information already available to it, namely, its ‘prior beliefs’, and uses this to derive predictions or hypotheses about what it doesn’t in fact know but what it infers or deems likely to be the case based on what ought to be the case in its world. Within the D-PE model, it is said that at every level of processing the system transacts this task by obsessively asking a single question, Is it important? of every would-be bit of information, i.e. pattern in noise, which comes by, and should the
system obtain an answer in the affirmative, then operations at that particular level select-
in the likely bit of pattern, pulling it out of the surrounding noise, and so making it available for possible further processing at the same level or levels higher-up.

As observed above, in the preceding case study, circumstances are such that Dennett (personality, body, brain) and his environment (physical, interpersonal, social, cultural) are operating more or less on a par. That is to say, there is no significant breach or deficit between oughts and world. It is by definition a normal state of affairs. However, as also observed, this is typically not so in cases of erotomaniac delusion. Specifically, in such cases, there is a marked disparity between what the system needs, by way of interpersonal relations or ‘companionship’ and what the world can deliver with regard to this criterion. So turning to the idea of prediction error processing, I propose that this disparity creates two problems for the system.

Problem one, concerns a lack of options, and it relates to the chance of finding a meaningful pattern in the environment. Essentially, here we have a living system which is obsessively searching its surroundings for a criterion, i.e. companionship of (one may venture) a specific sort, but which criterion is unavailable for selection. Basically, it is working to find something, a token, which just isn’t there.

Problem two, is an extension of the first problem, and it relates to risk. Namely, in this particular situation the system is faced with the option of selecting nothing or noise. However, it is reasonable to suppose that for a system which is driven to maintain its requirements and so is ‘epistemically hungry’ by design (e.g. Dennett 1991 [1993] pp. 16, 181 etc), nothing is not a tenable and certainly not a tolerable option. Indeed, under stress, due to mounting motivational pressures, here is a system which not only needs but really needs to find pattern x. It is not so much epistemically hungry as epistemically starving. Nevertheless, given the lack of options, there is no obvious x to find. So, for this reason, the system engages in some noisy or risky epistemic leaps: Is it important? ... Hmm, yeess, just perhaps, is the doubtful answer it tolerates in reply.

This is not far removed from the situation in which a living system faced with a genuine 50:50 stalemate between two equally favourable (or unfavourable) options must eventually opt one way or another, by design (see chapter three). And as the system’s opting in that case is a reasonable enough response to its particular (difficult) circumstances, so here the risky epistemic leap admitted by the system is a rational enough move. That is to say, the system is not acceding to epistemic risk because it is irresponsible, or addled, or deeply irrational. On the contrary, the system assents to
these jumps because it is a rational entity which is working precisely as it ought to work so to maintain the integrity of the entire body of works.

Still it might be objected that, no, the above behaviour is not actually rational, for in the continued absence of a suitable object the rational thing would be to relinquish the objective. And in response, I would say that to some extent, I agree, and I am sure that many systems do just this. However, I am also sure that many systems have access to alternative possibilities which some other systems lack. Plus at the level of processing which we are talking about here, the contents of the system’s options are simply not amenable to smart analytical appraisal. This is because the need for companionship, i.e. interpersonal relations, is a rudimentary drive towards a single objective, or rather towards an appropriate token of a single objective, even if the shape of that token becomes somewhat amended downstream. Yet, as discussed previously, the lower a pattern works in the system, so the less room the system and with it the subject has for manoeuvre. Which basically means that, ceteris paribus, in this sort of case, low fast risky epistemic leaps are going to be made.

(Slippery) Noise

So the situation is that, armed with expectation, the system inquires, ‘Is it important?’ of every bit of would-be pattern-in-noise that it finds, while simultaneously being disposed towards effecting increasingly risky epistemic leaps, until eventually the reply which bounces back from the unknown is a tentative ‘Hmm, yees, just perhaps.’ Now, from our vantage point looking in on all this, we (you and I) are in a position to know that by necessity, given the lack of available options, in this particular case the system has got it wrong. In other words, what has been selected-in as pattern is not actually pattern (or more correctly, the right sort of pattern). Instead, what has been selected-in is noise (or more correctly, the wrong sort of noise).

To explain. As discussed previously, noise is not just noise, noise is the medium of (all) pattern. Plus, as also discussed previously, this is not an irrational system. Accordingly, the noise which has been selected-in is not simply a random bit of informational junk. It is actually pattern, moreover, it is a pattern in which there must be something akin to whatever it is that the system is actually looking for. Nevertheless, it is a duplicitous or slippery bit of pattern which delivers an illusion to the system. In fact, one could say that, operating under stress leaves a system susceptible to being tricked.
Thus the situation is such that noise has entered the system, or more correctly that a parcel of the wrong sort noise has started moving around deep inside its informational pathways. So what happens next, in respect of the developing delusion? The answer is that, as regards the dynamics of information processing, pretty much the same as happened in the normal case of Dennett and the Molly B.

To explain. Think back to what occurs in the normal case when a specialist bit of the system asks ‘Is it important?’ and the answer it receives is, ‘Yes, perhaps’. This happens when there is sufficient correspondence between expectation and pattern to sustain the possibility that a prediction is right, but still too much error for the specialist function to be sure-enough re its interpretation, and which divergence causes tension, which upsets or excites the system, and which excitement supplies the energy for: (a) further examination at the same level of processing – perhaps following the acquisition of additional information into the system’s global belief set or via the system slightly amending its own hypotheses in light of new data; (b) foregrounding the pattern, at which point it becomes available to higher levels of processing, to which it presents as noise, and on account of which availability retesting at any or all of these other levels may or may not occur – but should retesting happen, then with each additional round of questions and answers, the system works to develop the brain’s own story about what is going on in the wider world.

Now think of the subject whose system is stressed but as yet exhibits no obvious symptomology. I have suggested that in a state of motivational agitation, such a system is prone to being tricked, i.e. making a grab at patterns which by necessity are not meaningful pattern as such, but which are really a slippery bit of noise. We can predict that in such a case, the schism between satisfaction re predictive success and uncertainty re error is going to be marked. Thus we can further predict that this situation is going to contribute substantially to the turmoil in an already agitated system of works.

Moreover, given all the excitement, so from the perspective of prediction error processing, we can further reason that the information which has been acquired re the slippery bit of pattern is going to get foregrounded to the nth, which means that in the blink of an eye information concerning the would-be-need-satiating-but-actually-false bit of pattern is going to be disseminated via noise across potentially every single level of processing in this need-hungry system. So that, if other conditions fail to hold, i.e.
certain constraints fail to be in place to check the progression of the duplicitous pattern, it is not difficult to see how an anomalous narrative could fairly rapidly take hold.

What is more, once the anomalous pattern starts to take hold, so its contents may become a self-fulfilling story, since if the system is actively on the hunt for patterns which support its predictions, this means that it is prone to finding precisely these patterns – and especially so given that it is already disposed to taking epistemic risks. Moreover, each ‘successful’ procurement of a subsequent bit of pattern effectively works as evidence to support the system’s (by now erring) hypotheses. So that step-by-step matters work to consolidate the abnormality, which progressively spreads and entrenches across the information processing network, and in so doing progressively effects an increasing dissociation between what for all the world really does seems to be the case to the brain and what really is the case in the world.

The explanation so far is that of a fundamentally normal system which essentially works as it ought, but which due to wider special or difficult circumstances, must open itself to risk. However, a consequence of opening itself to risk is that the system becomes vulnerable to noise. In particular, it becomes vulnerable to a slippery kind of noise which masquerades as viable pattern. And should a package of this (slippery) noise be selected and gain traction within the system’s information-sorting pathways, it can give rise to confusion, which in time can compound, to the end that the brain starts concluding that the world is a way which, objectively, it is not.

Therefore the D-PE account of delusion makes significant appeal to the idea of noise, and particularly to slippery noise, i.e. would-be pattern. So, it might be asked, in a case of erotomanic delusion, what is the crucial package of (slippery) noise a would-be pattern of? In response, I would say that with regard to any particular case of delusion it is impossible to pin-point an exact bit of duplicitous pattern which initially corrupts the system. However, we can speculate in general about the kinds of pattern which are likely to have a misleading effect, and also the channels through or times at which this is likely to happen, and identify certain characteristic features and consequences of such patterns taking hold.

So why in any particular case it is impossible to pin-point the exact bit of duplicitous pattern which corrupts the system? Consider a subject who has over time developed and so presented with the features characteristic of erotomanic delusion. If one probed her history to try and determine the precise root cause of her problem, literally all one would ever uncover would be competing likelihoods and shifting
patterns. It is the same as trying to specify precisely what leads Dennett to select the *Molly B* in preference to *Unicorn*. Really, who knows? Certainly not Dennett (Dennett 1978 [1981b]). And it boils down to the idea that the closer one looks (perhaps at different times and in different circumstances) the more patterns one finds, with no way of ever determining what the one decisive point of information, or in this case misinformation, actually was. Indeed, if there was one misleading bit of would-be pattern which set the problem off (e.g. Dennett 1991, 1991 [1993]; Dennett and Kinsbourne 1992).

Still, we can speculate in general about the kinds of pattern which are likely to have a confusing effect. All this requires is that we shift the level of the explanation to one which enables us to make broader observations relating to characteristic trends. And we can attempt this with some degree of confidence because we know that the human system operates by way of highly normative principles and that the problem is itself a pattern of norms, i.e. a recognised medical condition, which together allows us to derive generalisations (essentially meaningful stories) so as to explain certain regularities which are observed to occur, and then perhaps test our generalisations against the world. This is basically what happens when persons with relevant knowledges work to determine the trends, causes and contributory factors of *any* medical problem, and where the generalisations they derive might tell a useful story about, say, heredity, development, biology, behaviour etc., or as in this particular case, about certain conditions which are likely to be found in the subject’s environment, and especially in her interpersonal associations.

So thinking about a typical case of erotomanic delusion. Given the special and difficult circumstances re environmental and especially interpersonal associations involved, so from the point of view of a system which is fixatedly searching for patterns relating to tokens which (likely) promise to address a deficit and thereby satisfy a drive for companionship, so we may reason that in this sort of situation any bit of noisy pattern which even tangentially hints towards satisfying the requirement of companionship (e.g. friendship, romance, relationship) might erroneously find itself dragged into its informational workings. This should not be taken to mean that any such bit of pattern will necessarily be incorporated in any significant way into the system’s machineries. Indeed, for any bit of would-be pattern selected in error, one may suppose that tens or even hundreds or thousands of other noisy bits of pattern get disregarded as noise. As observed in chapter three, this is after all a truth-loving system (e.g. Dennett...
1987 [1989b, 1989c]). The point is simply that given the circumstances there is an increased likelihood that a misleading pattern, i.e. package of slippery noise, which somehow manages to meet certain looked-for prerequisites, could not only be selected-in as being of (likely) importance but could also (plausibly) gain a hold.

For instance, take the point that the object of an erotomanic delusion is often a notable individual, such as a celebrity. The likelihood is that the only contact that the subject has had with this celebrity is via images, say on a screen or in a photograph. So, thinking about a photograph, in just the same way that Dennett turned to magazines for information about sloops, so magazine photographs of celebrities contain patterns of information, often conveyed via close up images of faces, and often with the focus of the photograph set deep in the object’s eyes. Yet these are precisely the sorts of patterns which a lonesome subject’s need-hungry system is going to select-in as important, and in so doing release typical patterns of lightening fast responses in the roots of her biology which far outpace any opportunity for cognitive redress.

But why, one may ask, does cognition even need to be a factor? Why doesn’t the brain or biology deselect or even ignore the pattern on account of its being an image as opposed to a genuine need-satiating item, i.e. an actual person? The answer is because this is a very old system which is designed to find meaningful patterns, but which in its oldest lowest fastest hardwired machineries lacks the capacity to discriminate between competing tokens of likely patterns. For this the system needs recourse to its newer higher-order conceptual capacities, which afford it (and therein the subject) room for manoeuvre, but the trade-off for which is lengthier processing times, by which point base biology will have already started doing its thing and in so doing (potentially) affecting the corridors of informational processing all the way up.

And nor is this to suggest that a subject who develops erotomanic delusion has deliberately gone searching for celebrity photographs. She may have done so, certainly. Just as Dennett may have intended to visit the store to purchase a magazine about boats. However, personal intent is not necessarily a factor. For instance, the subject might just happen to be at the post office, queueing near the magazine rack, effectively minding her own business. Yet all the while her need-hungry system is skimming and scanning its surroundings for significant patterns, and in the process realising very many non-conscious behaviours. For instance, motor behaviours, such as adjustments to bodily position, or more subtle behaviours, such as rapid saccades of the eyes. And this continues until the system chances on a likely stimulus, say, a celebrity photograph on
the cover of a magazine. Whereupon information re that stimulus gets selected-in to the system’s informational machineries. Perhaps as it does so prompting further changes, and yet more responses. Yet, about all and any of these adjustments, the subject herself may be wholly unaware (e.g. Dennett 1991 [1993]).

And actually this could be the end of the story. That is, there may well be a subject who is information-wise functionally disposed towards certain patterns of information re celebrity pictures, but which dispositions are never fulfilled in any significant way. I suppose she could be said to have a functional (latent, unconscious) bias towards x. But a functional bias is in no way a psychotic disorder. Accordingly, the explanation needs to make the critical move from a system with corrupted informational processes, to a subject who develops certain problems which are recognisable as the psychiatric condition of erotomanic delusion.

A developing psychosis

The account to this point is that of a subject whose motivated system is vulnerable to the effects of noise, not because it is irrational, but because it is operating rationally, i.e. precisely as it ought to (by design) given the special or difficult circumstances that it (the system, subject) is negotiating at this time. One day, the subject, let’s call her Patsy, visits the post office to buy a stamp. She queues, quite by chance next to the magazine rack, where, on the cover of one of the glossy magazines, is a picture of a celebrity, we can call him Handsome. Perhaps Patsy already knows of Handsome or perhaps she does not. In any case, the information wrapped up in the picture of Handsome makes for a particularly slippery bit of noise, which successfully tricks Patsy’s need-hungry system (as opposed to tricking Patsy, herself) to infer that these are patterns of likely importance, that is to say, that they represent a promise for satisfying some of its most basic requirements. However, because of the circumstances, the system does not just attend to but gets very excited or agitated by the occurrence of these patterns, and which energy and misinformation has the effect of confusing it – and not just a little bit, which little then gets constrained or corrected downstream; and not merely to the end that the brain acquires (learns) a functional or dispositional bias towards what it now knows and anticipates re Handsome-type patterns; but in a way which without check compounds, markedly, across the hierarchical information-sorting system, so not only does Patsy come to entertain (i.e. assent to) the thought that he (Handsome) loves me, but she subscribes to a conviction that he loves me, is in contact with me, and (possibly) is
destined to be with me. And, moreover, she (explicitly, volitionally, intentionally) acts on this conviction to the detriment of her wellbeing, and possibly the wellbeing of Handsome or his family, to the end that Patsy receives a diagnosis of a major psychotic disorder and maybe a criminal conviction as well.

But why does this happen? That is, what is it that makes a fundamentally normal albeit deeply lonesome human being arrive at such a dreadful state of affairs? I have suggested that this outcome is in large part attributable to a certain type of noise, successfully tricking the need-hungry system into inferring that it (a bit of noise) is something it is not. However, a system is not the person, not really. And at the time that the mistake happens, while she is standing in the queue, Patsy could be unaware that anything out of the ordinary had occurred. That is to say, even if the photograph is located in Patsy’s perceptual field, it might be that nothing about it makes it into her experience of the world (e.g. Dennett 1991 [1993] pp. 333-8 ); or if Patsy does notice the photograph, it might be that nothing about it stands out; and/or as she leaves the post office after buying a stamp, she might have no memory of ever seeing the picture (e.g. pp. 115-26).

Nevertheless, we know that as regards the photograph significant changes are in fact happening deep inside Patsy. Specifically, her informational workings have judged that the would-be pattern is possibly important, and have done so because they are in part right (i.e. there is predictive success), but they are also wrong (i.e. there is error). And as in any normal case the tension between prediction and error causes systematic excitement, but in this case it is not mere excitement but stress, agitation, which triggers round upon round of rapid and wide-ranging retesting and a possible amending of hypotheses re Handsome-type patterns as the system urgently works to meet its own expectations. And the significance of this is that, via a series of normal operations, and entirely unknown to Patsy, her system is effectively pushing to repeat the misleading event.

The push to repeat happens because after achieving a (predictive) success of a kind, i.e. some promise of a hitherto unmet requirement being satisfied, an inferential system’s way forward must be to replicate the success. That is to say, across multiple stages, the system’s onus is to adapt so as to (re)find and (re)grab the pattern which successfully meets its expectations. This could mean changing its actual position in relation to a meaningful criterion and/or changing the patterns of information re that criterion which it deems important (e.g. Friston et al. 2006). Accordingly, Patsy revisits
the post office. The first time to post her letter. The next time maybe to buy another stamp. Yet if at each round of processing in which new information is added to the (Patsy’s) system its hypotheses are open to amendment, i.e. measured change, so with each round of processing a developing set of Handsome-related patterns could plausibly be gaining a deeper and wider hold.

Still, as the drive to repeat plays out across time, which could mean microseconds, seconds, minutes, hours, days or more, so it might be queried that longer timescales grant more room for cognition, so enabling Patsy to weigh-up what is happening re her new partiality for going to the post office at the end of the road. However, while in normal circumstances a relationship between time and a capacity for personal reflection may well be a factor, in Patsy’s case the matter is not clear cut. This is because, in a connected and stressed system, highly motivated patterns of (seeming) importance, would, as suggested previously, be disseminated bottom-end to top-end very rapidly, and which, reasonably, could facilitate not so much rational deliberation, as perhaps a variety of disjointed and emotive impressions or thoughts spontaneously arising, which is to say, from Patsy’s perspective suddenly ‘popping in’ to her mind.

Thus Patsy returns to the post office. Perhaps again and again and again. And with each return a deep expectation is (potentially) reinforced, which in each instance effects a burst of systematic thrill. Yet the thrill is hollow because a celebrity photograph is not a truly sustaining pattern, so what gets generated inside Patsy’s system is in some ways more akin to disturbance or upset than the thrill of true gain. And it is possible that her information-processing machineries might work this out, it is after all a ‘truth-loving’ system. This is to say, at any point, i.e. upon any repetition and/or round of prediction-error processing, it is possible that the negative impulses could win out. For instance, on a subsequent visit to the post office, as Patsy stands beside the magazine rack and her system is scanning its surroundings for pertinent patterns of information, the feedback it receives re the photograph of Handsome is not that of pattern but of error. At which point, various consequences could in theory arise.

First, the effect could be that the informational edifice or belief set constructed around the concept (of the photograph of Handsome) abruptly collapses, which would prompt inside Patsy’s workings something like a ‘change of heart’ – as contrasted to Patsy herself having a ‘change of mind’, which could possibly happen too. For instance, Patsy might have a sudden reversal of interest and decide that there is nothing of interest in the post office after all.
Second, the effect could be slower in that the system starts to devalue and so divert motivational energy away from that particular concept, in other words as regards any promise augured by Handsome Patsy’s system starts dreaming itself out.

Third, that particular error report could be ignored. That is to say, the overall belief set which the informational-system has erected around the concept (of the photograph of Handsome) could win out, whereupon the (as it happens truthful) error report is (wrongly) dismissed as noise. I would reason that the more need-driven Patsy’s system is and/or the more effort it has put into drafting its own story and hypotheses around the pertinent stimulus, the more likely it is that an ignore scenario would occur.

And perhaps it should be noted that here is the beginning of an explanation concerning the intractability of delusions, for the reason that even at this early stage, i.e. before content-fixations subsidising the pattern be loves me have coalesced across the brain’s narrative – and following on from this perhaps into Patsy’s observable behaviours, including her thoughts and experiences – there is a drift of basic informational patterns starting to accumulate in the informational pathways of the system around the concept (of the photograph of Handsome), and which although it is, as yet, only a set of predispositions, it may already be (on its way to becoming) a vital (energetic, aggressive) complex of (basic) belief which is opposed to, indeed is ‘logically’ resistant to, correction. However, as regards the developing psychosis, the question which now should be asked is how aware is Patsy about what is going on at this time? That is to say, as she returns to the post office, and peruses the magazine rack, does anything about her experience suggest to her that something is amiss?

As discussed in chapter two, the role of experience in accounting for delusions is a key point of theory for both the empiricist and relationist rationalist accounts – the empiricist maintaining that a prior (clinically) abnormal experience is the or a necessary explanatory factor in any case of delusion, while the relationist rationalist view is that (clinically) abnormal experience, by necessity, never assumes a role. In contrast, the D-PE standpoint is that since in a single connected system information is corralled in judgements and collapses into patterns of behaviour at different stages with different downstream effects – such as, motor and linguistic behaviours, thought or experience – so if (slippery) noise were introduced into the system, then there is no logical reason why in some cases of delusion anomalous experiences may certainly be a key explanatory factor, but in other cases this need not be so. And while, to some people,
this more flexible approach to delusions may be intuitively appealing, to others it may not be so.

For instance, Shitij Kapur (2003) argues for the idea that psychosis is a disorder of aberrant salience. Specifically, a subject undergoes an experience of an external item or an internal feeling or thought which is oddly salient, i.e. unusually foregrounded or having exaggerated importance in the perceptual field or mindset, and significantly the experience does not simply pass. Yet, consequently, the subject feels perplexity or anxiety re the experience, and so is motivated to explain what is happening to her and/or in her world. And it is during the process of trying to understand and assimilate the experiential oddity that a psychosis develops and takes hold (pp. 15-6).

Yet, from a theoretical perspective, if Kapur is correct, and Patsy has an experience of abnormal salience, then there would be nothing standing in the way of an empiricist one-factor theorist picking up Patsy’s case as it has been argued so far, flipping it around, and in so doing suggesting that when Patsy first stands in the queue, she experiences the photograph of Handsome as being unusually salient, and while her need-hungry system works to process the significance of the occurrence via its quite normal prediction error mechanisms, owing to the initial abnormal input, i.e. the oddly salient experience, the hypotheses it generates become increasingly anomalous, and as a result of which accumulative inaccuracy a case of psychosis eventually occurs. Hence, we would have a prediction error buttressed one-factor account of delusion, but with no clear way of telling whether actual experience (as per the empiricist standpoint) or seeming experience (as per the Dennettian standpoint) occupies the critical explanatory role.

However, if we allow for a moment the empiricist to take up the explanatory reins, the most obvious obstacle the empiricist thesis faces is that it is a clinical characteristic of erotomanic delusion, and especially of a case of primary monothematic erotomanic delusion, that the disorder presents with zero significant experiential feature, though in its later stages hallucinations are reported to sometimes occur. However, this presents a clear difficulty for any approach which demands that an anomalous experiential factor takes the central explanatory role.

In contrast, erotomanic delusion is a seemingly straightforward pattern of disorder to explain from the D-PE perspective of a located subject and the motivational-informational dynamics of a problem-solving system, because, as described above, the approach demands no necessary requirement for anything other than normal
experience as a developing psychosis takes hold. Nevertheless, if the clinical profile of the delusion warranted it, then the account could apply itself along different explanatory lines. For instance, say the occurrence of the photograph triggers a cocktail of fast biological changes in Patsy’s body and the downstream effects of this leads to her experiencing a sudden and inexplicable surge of suggestive feelings while standing minding her own business in the post office queue – which would be enough to nonplus anyone, I submit – and which new experiential information would then feed back into Patsy’s informational system, as is likely all lit up with importance and motivation, and so demanding to be explained therein. For sure, it could happen, and perhaps in certain cases of erotomanic delusion it does happen this way, maybe because, say, the subject has a background biography which includes having fostered an obsession with the idea of sex, or perhaps an aversion to the idea of sex, or perhaps there could be something ‘organic’ amiss, say, as regards her reproductive system. Even so, if the intention is to provide a model account of monothematic erotomanic delusion, then cases such as this, i.e. cases with atypical characteristics, while requiring an explanation, should not be conflated with the typical pattern of disorder.

Still perhaps it could be queried whether the one-factor empiricist could co-opt two-factor empiricist Max Coltheart’s idea of a significant experience qua functional and informational changes alone in the unconscious machineries of the brain (Coltheart et al. 2010), so that the first ‘delusion-specific’ experience that a subject such as Patsy has is when the fully formed delusional thought that he loves me enters her conscious mind. The explanation would obviate the problem the empiricist faces of accounting for delusions with zero anomalous experiential element, such as erotomanic delusion, within the empiricist framework. However, as I discussed in chapter two, the idea raises problems for Coltheart’s two-factor empiricist thesis, and the same problems hold for the one-factor theorist here. Namely, the one-factor empiricist would now be proposing two different theories of delusion, (one involving actual experience, and one not involving actual experience), for two different kinds of delusion (delusions with an experiential feature, and zero-experience delusions), which at the very least undermines the cohesiveness of the empiricist one-factor theorist’s approach to delusion. In response the empiricist might moot the idea of ‘unconscious experience’, i.e. actual changes in the machineries of experience which realise everything but an actual experience at the level of the personality. Yet, whatever one may think of this proposal, and Coltheart et al. explicitly rule it out (p. 264), there remains the problem that if functional and
informational changes alone are sufficient to effect a psychosis, then why take the additional step of awarding actual experience a role in any type of belief formation, delusional or otherwise. That is, to opt for the explanation suggested by Coltheart, not only throws into question the empiricist one-factor theory of delusion, but also the foundational empiricist conception of experience and its relationship to belief – which seems a high theoretical price to pay to accommodate erotomanic delusion.

Finally, with regard to the role of experience in the early stage of the disorder, does this mean that, according to a D-PE approach, Kapur’s observation re the role of abnormal salience of experience as the key contributing factor in cases of psychosis must be wrong? The answer is in part yes, but essentially no, not at all.

The way in which the two theories do not concord is that the D-PE model takes the position that abnormal saliences of experience are the or a key contributory factor only for (types of) cases in which the observed profile of the disorder warrants the factor of anomalous experience to hold, but not in all (types of) cases, i.e. not by necessity in every single one. For instance, regarding erotomanic delusion, which is a type of delusion in which abnormal experience is typically not a factor, I have worked to present a case in which Patsy, at an early stage of disorder, might not see (or remember seeing) the photograph of Handsome, but the information re the photograph is nonetheless awarded value by and so elicits changes in the non-conscious workings of her system – which changes may in time lead to subsequent downstream experiences of anomalous salience which perplex Patsy in the way that Kapur describes.

However, even in this case, the D-PE model is not entirely at odds with Kapur’s proposal, it just requires the attribution of salience to be shifted upstream. Specifically, we are not talking about salience at the level of experience for Patsy, but salience at the level of information in a need-starved system. That is to say, at pertinent levels of function certain would-be patterns of noisy information are selected-in as important and valuable, i.e. salient, and in so doing trigger excitement or agitation, which could result in the contents of these patterns (in one form or another), achieving purchase in the higher cognitive machineries of the living system, whereupon, if other conditions hold, anomalous downstream behaviours such as Patsy avowing that: He loves me, could occur. Basically, it is the same dynamic happening in the informational workings of a problem-solving system as Kapur might say would happen at the level of Patsy herself.

Furthermore, we could have a case in which anomalous experiential salience has a far more conspicuous explanatory role. For instance, take the alternative scenario
described previously of Patsy being suddenly overwhelmed by suggestive feelings. Or alternatively, consider the commonplace phenomenon of unexpectedly witnessing a face in the seemingly random patterns of an arbitrary object, as was discussed in chapter three on seeing Elvis. So we might imagine that one day circumstances (e.g. light, angle of observation, Patsy’s level of alertness etc.) are such that Patsy suddenly discerns what looks like Handsome’s face in the wallpaper. I would reason that, in conjunction with the wider special or difficult circumstances previously described, this experience could function as a particularly pernicious packet of duplicitous noise, simply because every time Patsy walked into the room it would be there, repeated. And once one has seen these sorts of illusions, although it is certainly not impossible that in many normal cases as circumstances change so the illusion simply vanishes, in other cases, if circumstances don’t change sufficiently, or perhaps if the system had a lot invested in that particular story, so it could be practically impossible to see the content of the illusion away. Furthermore, given the way in which the problem-solving system is held to function, so it is plausible that Patsy’s predictive mechanisms would at this stage be expecting to find (and thus more disposed to finding) Handsome-type patterns pretty much everywhere, and where each ostensible success would serve to feed yet more Handsome related ‘evidences’ back in. Nevertheless, as regards an explanation for a clinical case of erotomanic delusion, this remains a special story, and so not an explanation warranted for a typical case.

Moving on, the explanation now needs to shift from Patsy, a fundamentally normal human being, but with a stressed and struggling system which has begun to fabricate a belief set around a particular (duplicitous) object, i.e. Handsome, to Patsy, a stressed and struggling human being, who presents with a pattern of behaviour recognisable as psychosis, and specifically with the delusion that he loves me. In principle, this stage of explanation simply parallels the story of how Dennett makes up his mind to buy the Molly B, with one key difference, for whereas Dennett enjoys the benefit of (repeated, deep, rewarding) satisfaction, Patsy most certainly does not.

To explain. The account so far is that Patsy’s system, motivated to replicate a (seeming) success, adapts – which given the circumstances includes making changes at the level of Patsy herself – so as to make repeated trips to the post office magazine rack. Whereupon, with each visit, or purchase, or thought, the system (potentially) gets terribly excited as it (potentially) tests and acquires new information which (potentially) supplements and so modifies its belief set around the significant criterion, i.e.
Handsome. This is precisely the same story as that of Dennett gathering information about sloops. However, the critical difference between Pasty and Dennett is that, Dennett is located in an environment which is well-appointed enough to meet his requirements, and so with each replication, i.e. each round of processing, the message which Dennett’s system, and at some point which Dennett himself, receives back from ‘the universe’ (i.e. from the wider informational surroundings) is one of success or reward. That is to say, a little boost of gratification which at some level or another satisfies the system that a key need is being (or is on its way to being) met.

Patsy, however, is not so well-located, as by definition her environment contains a significant deficit, which her system has been pushed into taking epistemic risks in its struggle to redress, and in so doing it has, on this occasion, been deceived. That is to say, Patsy’s system has selected as important patterns relating to a criterion which is never going to effectively satisfy its basic requirement for companionship, and so with each replication, each round of processing, the message which Patsy’s system, and at some point which Patsy herself, receives back from ‘the universe’ is one of confusion, agitation, privation or upset. That is, despite the ostensible promise of satisfaction in the Handsome-related patterns; and maybe the Handsome-related glamour which might sort-of-satisfy Patsy’s imagination; and even though top-end changes can have effects which reach all the way down; it would take a very special story to convince Patsy’s biology, or indeed Pasty’s heart, that its deepest needs are truly being met.

So faced with this situation, what does Patsy’s problem-solving system do? Basically, the only thing it can do. It sets the machine in motion to yet again repeat the ‘success’. Why doesn’t it do something different? Because, as I reasoned a moment ago, whilst upon any repetition, it is possible that the negative (confounded, confused, agitated, distressed) impulses could in theory win out and the brain shout STOP!, given the complex of content which is building throughout Patsy’s system around the criterion of Handsome, as well as the dearth of viable alternatives in her world, it is by no means a straightforward prospect that the predictive mechanisms of her problem-solving system would or even by this stage could spontaneously opt in a different way.

And where is Pasty while all this is happening? Quite understandably, as time moves on, and trips to the post office add up, and photographs of Handsome litter her home and her thoughts, and confusion and agitation and upset start having evident downstream effects, Patsy is going to be more and more at sea in her world.
Thus unlike the normal case study in which the events of Dennett’s life run along pleasingly well, the case of Patsy tells of a subject who at the outset is negotiating difficult circumstances, but who, as events progress, finds herself struggling ever more against a swelling tide of circumstance and confusion and anxiety. And I propose that this is the point in the explanation where Kapur’s model of psychosis as abnormal experiential saliences intersects with the D-PE account of Patsy’s delusion, i.e. in a system (qua biology, human being) that is already working under significant stress.

As Kapur (2003) observes it is typical of (spontaneous, endogenous) psychosis that cases develop only gradually via a series of stages, which starts with the subject experiencing a state of ‘heightened awareness’ and agitation, i.e. ‘emotionality and anxiety’; followed by a drive to make sense of the situation, and which culminates in the delusion ‘crystallising’, whereupon the subject may start having hallucinations, which Kapur maintains are essentially cognitive occurrences, i.e. ‘exaggerated, amplified and extremely salient odd internal percepts’; and finally, if all the above leads to behaviours which are abnormal enough and public enough, then a clinical diagnosis, in this case of erotomanic delusion, may be received (pp. 15-16).

So, if we start with Kapur’s point re heightened awareness. As he observes, this is a ‘novel and perplexing state’ in which the subject experiences an ‘exaggerated importance’ re certain percepts and ideas which seem to stand out as salient in the perceptual field or her own thoughts, and importantly, these saliences don’t simply happen then die out, but they remain vitally alight in the foreground of awareness, e.g. ‘It was as if parts of my brain awoke, which had been dormant’, ‘My senses seemed alive. ... Things seemed clear-cut, I noticed things I had never noticed before’ (p. 15).

Yet, as has been discussed previously, once the internal tensions which are stressing the capacities of this living system start breaking through into observable behaviours, including experience, this is precisely what the D-PE model predicts a need-starved system which is intensely searching for meaningful patterns in the world is, at the subjective level of the personality, going to undergo.

Likewise, as regards the criteria of emotionality and anxiety, I would argue that this is precisely what has been discussed previously re the agitation and confusion and upset in the system, as the effects of these changes eventually break through into a subject’s, e.g. Patsy’s, observable behaviour. And again, the fact that the anomalous experiences don’t simply pass is predictable and has previously been considered at length, when it was suggested that, if matters are left to develop without check, then
Patsy’s system could quite straightforwardly work to lock itself into a corrupted and vicious cycle, and as a result of which the anomalies she experiences are very unlikely to simply go away.

So what about the subject’s drive to understand her situation? Again, from a D-PE perspective, this is what one would expect to happen, for the reason that this is a living system which is made (literally, by design) so as to be deeply driven to reach conclusion after conclusion about what is going on in its world and thereby to resolve the problem of what it ought to do next. It is the reason why Patsy’s system was pushed into epistemic risk-taking in the first place. And it is why we would expect a subject, whether they be Dennett or Patsy, to be motivated in precisely the same way as he or she works to make sense of his or her world – the difference being that while in Dennett’s case this happens effortlessly, with barely a ripple, in Patsy’s case, she will likely be alert to the unsettling effects of what is going on inside and around her as she, her system, works to try and navigate the increasingly turbulent conditions which appear to be flooding in.

Relatedly, Kapur asserts that delusion is a disorder of inferential logic (p. 15), meaning that in an attempt to work out what is happening, the subject applies what she already knows to make sense of what she doesn’t yet know, and an everyday example of which would be her concluding that the sun will rise in the morning for the reason that it has risen in the morning so many times before. However, the transition from known to unknown is precisely the principle at work in the D-PE model, whereby based on what it already knows the system projects prediction upon prediction onto the world, whereupon the world returns error report after error report as to how right or wrong a prediction is. Accordingly, move up through the hierarchy of functional stages, to the level where Dennett sits pondering what features he would like his boat to possess and in so doing drafts and redrafts his list of preferred properties, or Patsy sits worrying why certain pictures seem to have a deep meaning or significance which she can’t quite grasp, and in so doing comes up with possible ideas which she tests and re-tests against the photos of Handsome again and again and again. Likewise the same pattern occurs in the actual, tragic, but well-documented Tarasoff case (see chapter one), in which the subject secretly records his interactions with the object of his erotomanic delusion, Tatiana Tarasoff, then spends hour after hour scrutinising the recordings for clues or evidences re his ‘relationship’ with her (Lipson and Mills 1998).
Furthermore, the principle of the system/subject actively pursuing conclusions or resolutions or ‘evidences’ has two consequences both of which are relevant to the explanatory account. The first consequence relates to the confirmatory role of evidences. And from the standpoint of the subject per se, this is mentioned briefly by Kapur (pp. 15-6). However, it is a principle which has deep implications within a D-PE framework. Consider that said evidences come to light when armed with expectation the system, by way of its specialist structures, skims and scans its surroundings and hits upon pattern corresponding to its requirements, which not only rewards the system with a boost of satisfaction, but also has the effect of supplying not merely its problem-solving machineries, but also perhaps the subject, with reasons to support or strengthen the story which is taking shape inside and around her in her world. So, if we think in terms of Patsy. As she pours over photographs of Handsome, her system (she) is selecting-in information re confirmatory ‘proofs’ which are continually being incorporated into what it (and perhaps she) already knows, to the end that, if the narrative being constructed across her brain were to be prompted in the right way, e.g. if at a certain point one were to intrude and ask Patsy: What? or Why? x is happening, in all sincerity Patsy would likely be able to provide cogent enough ‘reasons’ to support what she was thinking or doing at that time. For instance, she might say: ‘But look! This links up with this and explains that, which in turn supports the other thing.’ And, I can say from practical experience that when faced with this sort of logic, it can be very hard to argue otherwise. In fact, unless it is approached with sensitivity, any attempt to point out the wrongness of the rationale can meet with a markedly unfavourable (emotional, linguistic, physical) response. Yet from a D-PE perspective such a response is not surprising given the amount of motivational (affective) energy tied up in this already precarious (confused, agitated) but deeply important (valuable) belief set.

The second consequence is that in the earlier stages of the psychosis with each bit of ‘proof’ acquired the agitation and confusion which is breaking this system would not be expected to lessen but might be expected to get worse. Kapur captures this point when he observes that the prodromal (pre-clinical or emerging) stage of a psychotic disorder can sometimes last for days or even years, during which time the subject continues ‘accumulating experiences of aberrant salience’ and so perpetuating the drive to ‘make sense’ of what is happening (p. 15). On the other hand, from a D-PE perspective, this situation has already been discussed at length in terms of the chronic cycle of agitation and confusion generated by duplicitous noise as a need-hungry system.
(subject) unceasingly works to negotiate its (her) world. However, as Kapur also observes, the tension usually stops when the obsessive ruminations terminate in a specific conclusion, i.e. when the delusion ‘crystallises’ or ‘psychotic insight’ happens (p. 15). It is a transition which is identical in kind to that which Dennett undergoes when he opts to buy (then subsequently changes his mind, and so opts not to buy) the Molly B.

It is a choice point that terminates a process of deliberation or consideration that is not apparently algorithmic, but at best rather heuristic. At some point, we just stop deliberating. We take one last look at the pros and cons, and leap. (Dennett 1978 [1981] p. 303.)

*Cognitive (psychotic) insight*

Psychotic insight for Patsy happens when she, at the level of the personality, participates in a ‘game changing behaviour’ by assenting to the judgement that: *He [Handsome] loves me*, and in the process she learns something about herself and her world (cf. Dennett 1978 [1981]).

As discussed previously, within this D-PE model, arriving at a specific conclusion is something which occurs by way of a prompt. The prompt may be pretty direct. For instance, Patsy might ask herself the question: What does this mean? and given the right set of circumstances this could be sufficient to collapse her brain’s narrative in such a way that she, Patsy, (eventually) assents to the idea that *He loves me*. Alternatively, the prompt could be related to the object. So, for instance, one day something about the photograph, or perhaps about a different photograph, or maybe something that Patsy hears Handsome say on TV works to initiate the crucial response. Furthermore, given all of the bottom-up motivation which is driving Patsy’s agitated system, so the crucial prompt could plausibly be something very trivial indeed – objectively speaking, even more unremarkable than the arrival of Dennett’s gas bill. So trivial, in fact, that Patsy fails to notice (or remember) that it happened. Yet the significance of this is that on reflection, perhaps, say, during a subsequent clinical consultation, Patsy could in all honesty claim that the ‘insight’ arrived or was conveyed to her from ‘out of the blue’. In other words, in line with the typical profile of erotomanic delusion, it could genuinely seem to Patsy as if Handsome had *made the first move*, i.e. initiated the relationship (Kelly *et al* 2000; Kennedy *et al* 2002).
Another point about Patsy’s cognitive insight is that, at first pass, it is unlikely to take the form of the neat grammatical expression *He loves me*. Instead, it is more plausible that her insight starts out as something more akin to a sense of significance or sureness, i.e. salience, regarding certain feelings, bits of thoughts, and/or bits of experience which over time get worked into a storyline which runs loosely, and only loosely, along the lines of *He loves me*. Why is this likely? The first reason is that this is the same principle which in the normal case study was illustrated via Joyce’s room, i.e. the excerpt from the novel *Finnegans Wake*. Namely, it is the point that polished grammatical concepts come together only gradually from out of a streaming disjointed jumble of informational ‘bits’. The second reason is that, given the depth of the need for companionship, it seems more likely that this delusion could coalesce not in the higher-order linguistic levels of the intellect, but nearer the system’s non-linguistic roots.

Thus we might imagine Patsy having a nagging feeling or impression that something of importance is going on concerning Handsome, and which (feeling) gets progressively run through multiple rounds of drafting and redrafting, until a story starts to develop which somehow satisfies Patsy’s personal requirements within the context of her particular circumstances re personal and social and cultural norms. For instance, if the environment in which Patsy is submerged elevates the notion of romantic love then perhaps it shouldn’t come as a surprise if the story which is fabricated through round upon round of non-conscious expectation and redrafting is one of quixotic relations with Handsome. Likewise, Kapur discerns the same point when he notes that delusions have psychodynamic themes and a cultural perspective which are relevant to the subject who imposes top-down explanations on their experiences of aberrant salience (Kapur 2003 p.15).

However, as regards this process, while in one sense the content of the delusional story clearly spirals outwards as a complex narrative of implausibility, impossibility or bizarreness. In another sense, the squeezing of Patsy’s initial (vague) insight into a (strict) linguistic structure is almost certainly going to involve an unrealistic narrowing of its content. It is the same principle as when Dennett was pressured into over-specifying his desire for beans (Dennett 1987 [1989b] p. 20). And the point holds regardless of whether the observer probing for conceptual edges is Patsy herself or someone else, such as a clinician taking her case notes.

PATSY: I have a feeling about him.
CLINICIAN: Can you tell me what that feeling is?

PATSY: Err, that he loves me?

Or, alternatively, when asked the question Patsy might engage is a long and disjointed account, from which the clinician moves to select-in what he or she judges to be a significant pattern, and which in the case notes might be recorded as:

Patsy maintains that the object of the delusion (Handsome) is in love with her.

However, in either case the proposition *He loves me/her* is likely to be over-specified or even off-beam with regard to the initial stimulus, i.e. Patsy’s cognitive insight. And there are a number of implications to this point, but if I may comment on four.

The first implication is that Patsy’s (diagnosis of) erotomanic delusion is not inevitable. That is, we are not treating Patsy’s delusion as a disorder which is hard-written in the structures of her brain, but as a downstream story, a fiction, which in light of certain special or difficult circumstances Patsy weaves around herself in a pretty typical but not inevitable way. Consider the point made in chapter three, and illustrated by the quote from Ted Hughes, that the same prompt at a different time or a different prompt at the same time may realise an entirely different pattern of outcome, and that with the passage of time those patterns are going to be increasingly intellectualised representations of the initial event. Nevertheless, that the story into which Patsy invests her energies happens to be *this* particular clinically specifiable narrative, I suggest, tells us something about the wider society and culture in which Patsy lives as well as the pathos which seems to frame her personal life.

The second implication is that, for Patsy, once the erotomanic-type narrative takes hold the story becomes a self-perpetuating myth. The informational dynamics behind this have previously been described in some depth in terms of a non-virtuous feedback loop which starts playing out across all the levels of informational processing throughout Patsy’s system. While at the level of the subject, Kapur observes what are plausibly the explicit effects of the process when he writes that:

> Once the patient arrives at an explanation, it provides an “insight relief” or a “psychotic insight” and *serves as a guiding cognitive scheme for further thoughts and actions.* (italics added, Kapur 2003 p. 15.)
The third point is that once the story achieves the status of a self-perpetuating myth, the subject could be expected to experience a degree of intellectual and emotional relief. This is because as the subject more-and-more lives inside her own story, so many of the conflicts in her belief set, and between her belief set and the world, are subsumed into the narrative, in this case concerning Handsome, which progressively becomes more-and-more objective or ‘real’. Thus we might expect there to be a lessening of the confusion and agitation which has so far driven the problem, both inside the subject’s workings and, downstream for the subject herself. Furthermore, it goes some way to explaining why it is that by the time the patient eventually arrives in a clinical setting and recounts the ‘facts’ of her case to a clinician, there may be no ambiguity to her mind as to who is in a relationship with whom. Even so, there remains the clear problem that this is a delusion, i.e. a personal self-perpetuating myth, so in a very real sense one would think that in any typical case there is always going to be some tension inside of this system, simply because the story is transparently false and so long as the subject remains living inside of it her needs remain fundamentally unmet. Hence, there is a very reasonable motivation to continue accruing evidences (Kapur 2003 pp. 15-6), both from the functional perspective of a problem solving system, but also from the perspective of a subject who, on the one hand, may enjoy the confirmatory ‘rewards’ that evidences deliver, but, on the other hand, may be trying very hard to maintain a very implausible story, not least because if holes were to appear in the narrative’s fabric, then in a very literal sense, it might seem to the subject that a significant (a meaningful, important, valuable) part of her world is in danger of falling away.

The fourth point is to simply spell out the consequence that, as Patsy’s delusion gains more and more purchase throughout her system, then given all of the brute motivation, systematic promise, accruing evidences, and emotional and cognitive satisfaction (and torment) which is sustaining the content of the delusion (as well as sustaining Patsy herself), so it becomes progressively unlikely that Patsy is going to, indeed simply could, change her mind. For sure, in time, Patsy (her system) might dream herself (itself) out. For sure, in time, small counter-evidences could accrue and in so doing effect a significant change. And for sure, the introduction of a personally significant game changing reason into Patsy’s life, such as arrest and detainment under The Mental Health Act (cf. Farnham et al. 1997), might quite speedily prompt in her a change of heart and/or mind. However, unlike Dennett who modestly reconsiders his
situation and in due course assents to the ‘game changing’ conclusion: *I don’t like the Molly B enough to buy her*, so it is very unlikely that after a bit of persuasion or simple reflection Patsy is going to renege on her narrative and re-reckon: *Oh whoops, I’ve changed my mind, Handsome doesn’t love me at all.*

**Hallucination**

Another feature to account for in Patsy’s developing psychosis is hallucination. According to Kapur, hallucinations are ‘by most accounts ... exaggerated, amplified, and aberrantly recognised internal percepts’ (Kapur 2003 p. 15). They are also a fairly advanced feature of a developing psychosis in that they occur following the cognitive insight and so after the theme of the delusion has taken hold (pp. 15-6). This description of hallucinations accords with the profile of hallucinations as they are said to occur in cases of erotomanic delusion, in so much as, if hallucination features at all, then it occurs at a later stage of the disorder, and is typically vague or dreamy and somatic in nature with a content which is clearly related to the theme of the delusion (e.g. Fitzgerald and Seeman 2002; Kennedy *et al.* 2002).

From a D-PE perspective, experiences such as these can be straightforwardly accommodated. Consider that here is a system in which changes at the top-end, i.e. abstracted thoughts, can influence changes all the way down into biology. And changes at the bottom-end, i.e. biology, can influence changes all the way up into abstracted thought. And between top-end and bottom-end are multiple levels in which patterns of information relating to perceptual-type contents are drafted. So we only need to assume that some of these bodily and intellectual changes realise patterns of likelihood which the system itself judges to be important enough to warrant attention, upon which, if other conditions hold, for the subject it could seem that a perception of $x$ is the case.

Moreover, it is likely that at first pass the content of the hallucination will be weak. Yet in a process mirroring the way in which Patsy arrives at her cognitive insight, so a story may start to develop around the anomalous experience. The content of this story will be partly biological, partly intellectual, and partly experiential. And in keeping with the idea of a corrupted or non-virtuous feedback circle, so as patterns concerning bodily changes, thoughts and experiences are repeatedly (re)selected-in as new information, so the system may start to support a self-perpetuating myth. Whereupon, the contents of the hallucination could get progressively more narrative and objective (Dennett 1991 [1993] p. 7). For instance, the subject may start to have a dreamlike
impression that the object of her delusion is in her bed at night (Fitzgerald and Seeman 2002).

A question which might be asked about this is why don’t all patients who present with erotomanic delusion report having hallucinations? The answer is that I don’t know. Maybe because they (systems) don’t all invest the same amount of energy into the more experiential aspects of the erotomanic-type narrative. Or maybe they do invest energy into all parts of the narrative, but for one reason or another those changes at the level of experience don’t make it into awareness, or aren’t retained in awareness. Or maybe they are, but the subject gives them less credence, or is less surprised, or less perturbed by them. Or perhaps they simply don’t recognise the experiences as hallucinations, but just pigeonhole them as normal bodily responses to, say, thinking about x.

**Diagnosis**

As Kapur observes, ‘So long as all these events (delusions and hallucinations) remain private affairs, they are not an illness by society’s standards’ (Kapur 2003 p. 16). That is to say, so long as Patsy successfully keeps information concerning her feelings, thoughts and experiences about Handsome to herself, then no external observer is going to look at her patterns of behaviour and conclude that something (of significance) is amiss.

It is an idea which coincides nicely with the continuum hypothesis of psychosis, which is the view that the features which we, in our society, have come to recognise as psychosis, i.e. delusions, hallucinations and attendant odd behaviours, are not features exclusive to psychiatric illness, but they occur on a continuum across all levels of the population, normal (healthy) or otherwise, with only the more pronounced or difficult cases being admitted into the clinical sphere (e.g. Johns and van Os 2002; Verdoux and van Os 2002; Rossler et al. 2007).

It should also be noted that from the D-PE perspective any observer of Patsy’s behaviour is going to be doing so from a particular standpoint. That is to say, from within a specific cultural, social and personal environment and via knowledge relating to his or her personal biases, biography, background gen etc. And from this standpoint, selecting-in one possible pattern from the infinitude of patterns potentially available, and thus constructing a (or rather, another) meaningful story of his or her own.

Accordingly, the clinician observes Patsy, and so discerns a meaningful (clinically significant) pattern in her actual behaviours, which in the case notes might be recorded
as: *Patsy maintains that the object of the delusion (Handsome) is in love with her.* Thus a diagnosis of erotomanic delusion could be made. Whereupon, this diagnosis, as a pattern of information, passes into the body of environmental noise, and from there (potentially) feeds back into the next round of interactions for Patsy, the clinician, and whomsoever else has a stake in these events.

*Some final queries*

In opposition to the account it might be queried that if we assume that there are very many lonely people in the world, then why isn’t erotomanic delusion far more prevalent? Which is a point closely tied with the related question of why don’t all early indications of erotomanic type disorder develop into being a clinical case? In response, I would say that the answer is pretty much the same as the answer to the query in chapter three as to why doesn’t everyone want a boat? That is, perhaps erotomanic type thoughts are actually very prevalent throughout the population, but for one or another of the reasons mentioned above they don’t come to the attention of significant others. Or maybe erotomanic type ideas could be prevalent, but aren’t, because different people have different capacities, backgrounds, opportunities, inclinations etc., and any or all of which factors could affect the development of the relevant pattern of behaviour.

Finally, it might be queried, certainly this is a wide account, but couldn’t we make it narrower, maybe focus just on the brain, or perhaps even just on prediction-error reporting? Which closely ties in to the question of whether I am suggesting that a wide psychodynamic explanation works in every single case? In response, I refer back to the idea of different explanatory stances (e.g. Dennett 1987 [1989b], 1991). And I suggest that, for sure, one could adopt a physical or purely design stance towards the problem of delusion, and in so doing focus solely on biological criteria or prediction error theory alone. However, two consequences of this would be that: first, one would miss out on an entire strata of (potentially very) meaningful patterns relating to human beings and the hows and whys of their interactions with each other and their environments (e.g. Dennett 1987 [1989b] pp. 25-9); and second, as the above discussion of erotomanic delusion has hopefully highlighted, in some cases of delusion the root of the problem is likely not located in a subject’s biology, but in the subject’s relationship to certain wider factors, such as background, or location, or his or her interpersonal connections to others in the world (e.g. pp. 17-20). However, I am positively not suggesting that biological disorder is never a factor, since in some cases of delusion
biological deficits of one kind or another certainly seem to correlate with the delusion at point. So it is on this note that I turn to the second abnormal case study, which is Capgras delusion and the cases of Mme M and Mr A.

Capgras Delusion

As I outlined in some detail in chapter one, Capgras delusion is characterised by a subject maintaining the implausible conviction that a person known to them has been replaced by an ‘imposter’ of some sort. There seems to be no clear trend re the demographic characteristics of persons who develop the disorder. And traditionally explanations for Capgras delusion have fallen into two camps, psychodynamic (i.e. based in motivation and information) versus organic (i.e. based in a neuropathological deficit) (Ellis and de Pauw 1994).

So the question is what can the D-PE model which has been developed say about Capgras delusion? And to answer the question, I propose approaching the delusion in two ways.

First, I take a case which seems to lean towards a psychodynamic explanation, namely that of Mme M, who developed a pertinent pattern of symptoms following the death of her twin sons and the breakdown of her marriage (see chapter one, and Bentall 2003 pp. 309-310), and I propose that, on the face of things at least, the D-PE model developed in the previous section is well capable of applying itself to this case. That is to say, on account of the very special and difficult circumstances in which Mme M finds herself, her information processing system is pushed into working under stress, which leaves it vulnerable to noise, to the end that certain noisy patterns gain traction within her belief set, and subsequently in her behaviours, and which behaviours include a pattern which comes to be differentiated as Capgras delusion.

Second, is the case of the patient I refer to as Mr A who sustains serious head injuries which seem to precede in a significant way his subsequent diagnosis of Capgras delusion (see chapter one, and Alexander et al. 1979). Again, I propose that a D-PE approach can meet the conditions of the case for the reason that the effect on the system of the biological trauma sustained, in conjunction with the immediate and subsequent psychological circumstances it/Mr A has to negotiate, are sufficient not only to put an otherwise normal enough information processing system under stress, but also
to work as a direct channel by which noise can enter the system and (given other conditions) take a hold.

If this latter explanation holds, however, it means that with regard to Capgras delusion, a D-PE approach largely erases the uncompromising either-or line which is traditionally drawn between an organic versus a psychodynamic explanatory approach.

So taking each of the case studies in turn.

The case of Mme M
If we apply the D-PE model to the case of Mme M, in many ways the explanation parallels the explanation re Patsy and erotomamic delusion, with one clear difference. The difference is that while in Patsy’s case the dynamic described as motivating changes from the level of biology upwards is that of a need satisfying positive reward, (i.e. attraction towards or the energy behind the disposition to ‘grab’ at a stimulus, qua inferred boon), in Mme M’s case of Capgras delusion, it seems more likely that the primary driving force is not that of positive reward, but of negative aversion (i.e. satisfying a need by propelling the system away from a stimulus, qua inferred harm).

Yes, it is logically possible that the explanatory dynamic could actually be that of an ‘ignore’ response (i.e. an absence of motivation due the system disregarding the stimulus by failing to distinguish it as important). And for sure an absence of motivational affect re the object of the delusion is a characteristic feature of some cases of Capgras delusion, including the ‘organic’ pathology case of Mr A which I look at next. However, the dynamics of an ignore response do not sit well with the idea previously discussed re motivational salience. Moreover, given the details of Mme M’s case, I maintain that it is realistic that negative motivation is the prevalent explanatory dynamic in this case.

So turning to Mme M. On the one hand, Mme M is a system with biology at its bottom-end, intelligence at its top-end, and within which highly connected framework information moves around. On the other hand, Mme M is a human being who is clearly facing special and difficult circumstances which are more than sufficient to stress her system from its roots to its top. Thus as she stands watching the coffin of her young son being lowered into the ground and becomes convinced that the child inside the coffin is in fact not her own, we can speculate that quite understandably Mme M, her system, is beside itself re the significance and terribleness of what it is witnessing. That is to say, facing these patterns, the question, Is it important? which is obsessively being asked across her system is likely to receive back in answer a resounding YES! There is
no ambiguity here. However, the value that the predictive mechanisms assign to these very important patterns is not that of positive reward, as was coded in Patsy’s system when it detected the pattern of Handsome, but a negative aversive value, which instructs Mme M (her body and mind) to recoil from the stimulus.

So what happens next? I don’t actually know, but we can reason that perhaps Mme M is literally, bodily, motivated, and so runs from the graveside. Or perhaps her body stays where it is, and only her mind ‘exits’, and she swoons. Or perhaps, neither of these things (can) happen, but nevertheless her system is driven to find another way out. For instance, given the confusion and affect which must be tearing this system apart, so the energies generated are likely to have two significant effects. The first effect is that the storm of energy and confusion is likely to generate noise. That is to say, it is not implausible that Mme M’s informational belief sets are literally being torn apart. Putting it bluntly, she, her mind, is breaking up. The second effect is that, if in normal circumstances a system is motivated all the way up by what Dennett calls affective ‘editorial spin’ (Dennett 1991 [1993] p. 181), then in Mme M’s case it is not so much spin as editorial hysteria, and I use the term ‘hysteria’ non-pejoratively simply to reflect the stress, confusion and emotional trauma which I have suggested is quite literally tearing Mme M apart. Yet what this means is that the command to retract, renounce, withdraw pretty much explodes outwards and upwards across the system’s informational pathways. So that cognitively Mme M might start with the knowledge that her child is in the coffin, but from the roots of her biology the imperative message is that, no, withdraw, renounce, retract (from that sealed box). Round after round after round. So does cognition win out, or does biology, mother biology, win out? In Mme M’s case we know what the answer is, namely she starts to think that It is not my child (in that box).

Moreover, if we take the factor of all the confusion and noise and energy, then it is easy to see how the belief-set around the stimulus of the child could start pushing outwards. So that, with nothing to constrain it (and in this case, I wouldn’t like to suggest what the constrains could be, simply because I fear that I can’t really conceive such things without sounding trite), a succession of misidentifications ensue. It starts with the husband. Then pushes out, wider and wider. Hence, in Mme M’s case, the delusion does not remain straightforwardly monothematic, but develops to become a more extensive and complex psychiatric problem. Nevertheless, the dominant pattern throughout remains recognisable as that which is known as Capgras delusion today. So,
thinking back, could Joseph Capgras have observed Mme M’s behaviour and discerned a different pattern to which he subsequently gave his name? Yes, in all probability he could have. Yet given the meaningfulness of the pattern which Capgras did derive, so we can take it that he probably identified a pattern which he ought to have identified (by design).

An objection to this account of Mme M’s case of Capgras delusion might be that it is a neat little story, but nothing more. To which I would respond, of course it’s a story, what else could it possibly be? Moreover, it is a story with parts and chapters across many dimensions. So that if someone with the appropriate skills were to step in to assist Mme M, then there are many different entry points into the narrative through which they might begin. Yet I in no way wish to suggest that this would be an easy task. Not in a case in which such an emotionally charged belief set has taken such a thorough and complicated hold.

*The case of Mr A*

The D-PE account of delusion as it has been developed through the previous case studies tells of a fundamentally normal (healthy) living system, caught up in special or difficult circumstances, which result in its straining to work under operational stress. It is a situation which renders the system vulnerable to noise. Yet should noise gain a hold, then in the absence of a suitable constraint, it (noise) may secure a deepening role the (informational) pathways of the system. So that, in time, the narrative which is continually playing out in the subject’s brain and behaviour (including experience) could twist in a significant way. At which point, the subject effectively starts living inside a personal self-perpetuating myth. And if the observable indicators of this myth start to jar sufficiently and in a particular way with the pervasive environmental norms within which the subject must by necessity act and interact, so then and only then might the subject be awarded a diagnosis of a recognised clinical disorder, such as erotomanic delusion or Capgras syndrome.

So thinking about the cases of delusion which have been considered so far. In both of them, the special or difficult circumstances which initially placed the system under stress have been external to the subject’s body, and specifically to the brain. That is they have been *wide* explanatory criteria. In Patsy’s case, they related to her social and particularly interpersonal environment. In Mme M’s case, they started with the deaths and burial of her children. However, if we now turn to this case study in which the
subject, Mr A, developed Capgras delusion following a road traffic accident in which he sustained serious head injuries, then while there are certainly wide explanatory factors at work in his disorder, and I’ll come back to these in a moment, the special and difficult circumstances which are of particular interest to us at this point are narrower in the sense that they concern actual damages to Mr A’s brain.

A human brain is a staggeringly complex organ and the effects of damages to a brain can likewise be hugely complex. However, from the point of view of a D-PE thesis which provides description at the level of how an intelligent human system ought to behave (by design), so it is possible to condense the complex effects of actual damage, into the far simpler observation that the ‘organic’ damages to Mr A’s brain could function as a gateway into the system for noise, and specifically for noise which could be of a very special and difficult kind.

To understand what makes the sort of noise which could feature in this kind of case so particularly difficult there are three basic points to consider. The first point is that depending on the precise nature of the damages in question, i.e. the structures and attendant functions involved, the noise the damage generates could begin very deep inside the workings of the biological system, that is to say far outside the remit of considered control, which effectively renders the entire working system, bottom to top, open to its effects. The second point is that, to some extent the introduction into the system of this noise is going to be chronic, i.e. there is going to be more and more of the same kind of error introduced into the system until healing of one sort or another takes place. And the third point is that, potentially, the nature of this noise, its content, could be very odd indeed, for the simple reason that changes to the brain can lead to some very unusual effects, as anyone who has experienced migraine or concussion may attest.

Given the nature of these difficulties, it might be thought it is almost inevitable that downstream problems such as delusions are going to develop. However, within a D-PE framework, this is by no means a given outcome, and in fact the opposite expectation holds. This is because, certainly there is noise, error, oddity, which the otherwise normal system with its obsessive drive for survival is going to have to work (possibly very hard) to overcome. However, since the entire rationale of the informational system’s machineries is to minimise error and maximise pattern, so the expectation must be that oddity will be selected-out, i.e. ignored or corrected, and pattern will win out. That is to say, the crux of the D-PE thesis is that each one of us
enjoys a fundamentally marvellous truth-loving problem-solving arrangement of tissues and operations, i.e. a living human system. Nevertheless, it has been proposed that in some circumstances even an entirely undamaged system’s machineries can be tricked or overwhelmed. And I suggest that in the case of Mr A, whose brain is contending with a narrow set of special and difficult circumstances as well as those in the wider environment, this is a different instance in which understandable enough mistakes get made, and so are woven into the ongoing narrative, and which (without constraint) compound.

Accordingly, a D-PE account of Mr A’s case could be outlined as follows. To begin we have Mr A, who is a fundamentally normal (healthy) subject. Yes, there are some problems with his brain, but these are special considerations for a subject whose system is nevertheless working hard to function as it ought, and, one might add, is almost wholly succeeding in doing just this. All the same Mr A (or rather, his system) is in a situation of needing to negotiate certain special and difficult circumstances. Specifically, although prior to the accident Mr A seemed to enjoy a favourable enough life including a good relationship with his wife and children, following the accident he (his system) finds himself (itself) needing to navigate a knot of special and difficult factors relating not only to the actual effects of the brain damage, but also to the fact of his involvement in a serious accident, plus the fact of his injuries, plus the fact of his (long) hospitalisation and rehabilitation, plus his ongoing problems concerning motivation, memory, recognition, and, significantly to issues relating to changes in family roles. By any measure this amounts to a very substantial complex of associated problems, which in very many cases would be sufficient to place the living system under operational stress. Moreover, add to it the factor of informational noise of a kind which is chronic, possibly deep and perhaps weirdly atypical, and which is thus well placed to insidiously corrupt the informational pathways of Mr A’s system. The end result is a man whose system is motivated by need, yet addled by tension and noise, so that an atypical narrative starts to develop, which without constraint (e.g. the swift restoration of normal-enough brain function) in due course leads to Mr A effectively living inside a personal self-fulfilling myth in which for all the world it seems to him that: *Something about my wife and family is not as it ought to be, something about them has changed.*

Clearly this is the same model of explanation as given for Patsy and Mme M, meaning that from a D-PE perspective there is no need to change the basic explanatory approach to accommodate the factor of Mr A’s ‘organic pathology’. Even so, there are
several points of note which are more specific to Mr A’s, as opposed to Patsy’s or Mme M’s, case.

The first point of note concerns the role of motivation. It has been suggested that in the cases of Patsy and Mme M circumstances coalesce to realise an exaggerated motivational drive towards or away from a significant stimulus which in each case results in mounting agitation and confusion, but in Mr A’s case this pattern doesn’t seem to fit. Indeed, we are told that one of Mr A’s core problems is an absence of motivation. So what could be going on? In truth, from an armchair perspective I think it is impossible to say. Nevertheless, it could be reasoned that, given the way in which in normal circumstances motivational changes start deep in biology and then energise informational and behavioural changes all the way up and out, so the root of Mr A’s problem could plausibly be deep in the motivational machineries of his system, perhaps in the upstream neural mechanisms which control the initial coding of patterns as important or valuable (or not) in the brain. That is, for one reason or another, and feasibly due to actual damages, Mr A’s brain could be registering an ignore response more than it should. Alternatively, the problem could have its roots in the way in which motivational energy is generated and disseminated. That is, something is effectively interrupting the pathways in Mr A’s brain and body along which energy and information typically shift. So, in this situation, the early coding would prompt the right messages, but those messages would fail to arrive at the intended destination. And again actual damages or chemical changes could be suspect causes of this. Then again, the problem could be a result of something like depression imposing a dampening effect which could, in theory, reach all the way down. And yet another possibility is that in Mr A’s case a combination of these factors could be concurrently at work. However, re any of these possibilities, to an uninformed observer, Mr A’s outward behaviours could seem more-or-less identical, even though beneath outward appearances the causal pathways by which his ‘absence of motivation’ is realised could be quite different indeed.

A second point of note is the question of why the object of the Capgras delusion is Mr A’s wife and children? However, taking the shortfall of motivation to be the critical factor in Mr A’s delusion, then it may simply be that because his wife and children are most important to Mr A, so the incongruities of motivation could cluster, compound and hence in time become (more) noticeable around them. However, I doubt that this explanation is the rule. Consider that a different dynamic might be that if Mr A’s love for his family were strong enough to somehow ‘bulldoze’ information
about them (and only them) through whatever it is that is causing the motivational deficit while leaving everything else in his life behind the blockage, then Mr A would experience his family as being lit up by positive motivation in sharp contrast to everything else in his life which would effectively embody a dearth of affect. I don’t know if this actually is a pattern of a recognised mental condition. I can easily imagine that it is. However, perhaps because of the nature of the damage to Mr A’s brain, such ‘bulldozing’ would be impossible in his particular situation. And, in any case, positive motivational salience is simply not a feature in Mr A’s case. Finally, another possible scenario would be if the object of the Capgras delusion were someone whom the subject did not know particularly well. A case in point would that of a man whose delusion pertained to a particular nurse (Ellis and de Pauw 1994 pp. 323-27). Why this nurse? All things being equal, I can only speculate that she is someone who just happened to be present enough in circumstances which prompted enough of a response by the subject’s system at that particular time. It’s the same as asking why Dennett chose the Molly B over Unicorn. At the end of the day, the answer has to be, who knows?

A third point of note concerns what it is that prompts Mr A to make up his mind that his family has been replaced by imposters? To which the answer is, Nothing. For, almost certainly, Mr A didn’t make up his mind to this outcome, at least not at first pass. And especially so given the motivational deficit which is a central factor in Mr A’s case study, because language takes effort, and it would require significant motivational expenditure to supply the energy to bring into the world a concept as complex as: 

*Even though these people seem identical to my wife and children, they are not my wife and children, (therefore) they have replaced my wife and children, (therefore) they are imposters.*

To illustrate the point, think back to the *Finnegans Wake* example of Joyce’s room and the litter of rudimentary language fragments which only via time and effort (might possibly) come together into conventional, meaningful, expressions; or think about what happens when Dennett is pushed to over-specifying re his desire for beans, i.e. he has to grope for words which then don’t really fit; and think of Patsy anxiously gathering and testing sentences as she works to make sense of her perplexing thoughts, feelings and experiences. This is not to suggest that the notion of Mr A’s family being imposters couldn’t coalesce in time. Indeed, it did coalesce, not least because I’m writing about it now. However, as for when it first happened, at the very most, all we can say is that it came about only after countless rounds of hidden and explicit predictive analyses and bottom-up and top-down redraftings. Nevertheless, once the sentence did arrive in
the observable world, and was fixed down as securely as it ever could be fixed down in conversation and/or writing, say when it was first recorded in the case notes, then in all sincerity various parties might swear that this exact proposition had been a feature in Mr A’s case all along. At least, they might until someone reads back through the case history and discovers that this isn’t actually how it happened, or more correctly how it seems to have happened at the time. This inability to specify the time at which the definitive sentence appeared is directly comparable to the example of trying to pin down the definitive time at which Dennett’s list of favoured properties is complete. It never is. At any point in time it is open to reworking.

Perceptual Delusional Bicephaly

The case study by Ames (1984) which I outlined in chapter one describes events relating to a subject, whom I refer to as Mr P, surrounding his admission to hospital with serious gunshot injuries to his head. From what Ames writes it is clear that at the time of admission Mr P is already caught up in a storm of special and extremely difficult circumstances. He suffers a longstanding major psychiatric disorder (schizophrenia), for which he has had repeated hospitalisations. He was the driver of a vehicle involved in an accident in which his wife died. He has problems with alcohol consumption. He suffers delusions and is tormented by very strong multimodal hallucinations. And the confusion and agitation associated with these factors is clearly pulling him apart. Given this background, I suggest that talking through the impact that agitation and confusion have on Mr P (his system) would merely be repeating with emphasis what has already been said re Patsy, Mme M and Mr A. Thus I shall take this part of the account as given, and instead concentrate on the most striking feature of Mr P’s case, which is Mr P’s unswerving conviction that he has, clearly perceives, is in conversation with, indeed, is the victim of intimidation by, an intellectually autonomous second head, which eventually worn out by its coercions Mr P attempts to shoot dead.

So thinking about hallucination with regard to psychoses. According to Kapur, hallucinations are an advanced feature of a psychotic disorder, which, if they occur at all, do so after the content of a delusion has crystallised, i.e. assumed a particular form, and so occupies a place in and exerts a progressive influence on the subject’s ongoing cognitive framework; and where what the subject takes to be a hallucination is in fact a highly salient cognitive artefact from their own stressed cognitive processings, i.e. an
abnormal or unusual delusion-related imagining or thought, which the subject mistakenly fails to recognise as such (Kapur 2003 pp. 15-6). That is to say, ‘Hallucinations by most accounts are exaggerated, amplified, and aberrantly recognized internal percepts’ (p. 15). And I think that broadly speaking, Dennett would not disagree with this view.

The reason that I think that Dennett would (only) broadly agree with Kapur’s view of hallucination is that, as outlined in chapter one and discussed in chapter three re the case study of Dennett and Molly B, Dennett places all experience within the cognitive sphere, and so not just the anomalous experiences associated with hallucination. Contrast this to the standard construal of experience as actual discrete and fundamentally different to cognition. Unsurprisingly, Dennett’s reappraisal of the nature of experience is not without its implications. However, concentrating on delusions and anomalous experiences in relation to delusion, one encouraging implication is that Dennett’s approach to experience can readily account for cases of delusion across the clinical dataset. That is to say, unlike the prevailing theories of delusion, a D-PE approach to delusion encounters no theoretical glitch between cases which typically feature zero, weak, and (very) strong experiential features.

To explain. If we say that hallucination is a perceptual experience which occurs in the absence of a significant appreciable stimulus. Then as I argued at length in chapter two, the standard view of experience fails, because whether one approaches the problem of delusion via an empiricist model, resting on a causal theory of experience, or a rationalist model, resting on a relationist approach to perceptual experience, the role awarded to actual unified phenomenal experience, i.e. experience as it is standardly construed, runs into a tangle of difficulties and so fails to meet the data set of clinical delusions. At the very least, as Dennett observes, following the standard way of thinking about experience, while (very) weak hallucinations are theoretically possible, i.e. in certain circumstances a subject could experience, say, feelings such a tickles or transient pains or maybe flashes of light in the absence of the brain engaging with a significant stimulus (Dennett 1991 [1993] pp. 8-9), it is nevertheless the case that (very) strong hallucinations of the kind associated with psychosis are impossible, in the sense that it would be impossible for a brain to generate these kinds of very strong, multimodal, interactive experiences by ‘pulling its own wires’ in the absence of a meaningful actual stimulus, i.e. working abnormally on its own accord (pp. 3-10).
In contrast Dennett proposes an explanation re cases of (very) strong hallucination which starts with the idea of noise. Essentially, in certain circumstances, a system acquires a ‘lowered threshold for noise’ (p. 13), which is comparable to saying that in certain circumstances a problem-solving system opens its predictive mechanisms to epistemic risk. However, given this ‘lowered threshold for noise’, so a relatively trivial event, such as a fleeting perception or weakly anomalous sensation, could be awarded by the system more significance than it ought to be awarded. Whereupon, via a series of rapid-fire changes and multiple rounds redrafting, an anomalous narrative, qua ‘detailed fantasy’ (p. 7), develops in which related conceptions – including of a kind which inform the system via the channel of experience – become progressively stronger, i.e. more and more ‘objective’ (p. 13), which again is comparable to the idea of Patsy et al. effectively beginning to live inside a personal self-perpetuating myth.

Thus thinking about the case of Mr P. From the perspective of the D-PE model developed here, we can take it that within the context of the turmoil which is stressing Mr P’s system ‘something’ occurs in his (or, a particular) local environment. As Dennett observes, this something need not be anything particularly anomalous. And, certainly, as I argued re Patsy, it could be something which is essentially commonplace, such as an image in a magazine, but something about it, a pattern within it, is sufficient to excite (even more) an already over-agitated and over-confused informational system. So that, via round after round of rapid-fire functional changes, by way of which informational patterns are drafted and redrafted in a deeply impersonal process, given the continued absence of an effective constraint, an increasingly odd narrative starts coalescing across the (Mr P’s) brain.

Some layers of the brain’s narrative will be fast, other layers will be slow, and yet others will involve experiential judgements being manufactured then amended in the system. So that, across time and taken together a scrappy yet rich brain story develops. And part of this story, so we must assume, is potentially accessible to Mr P via the channels of experience and thought. So that at some point should he (his system) be prompted in the right way, say via a direct question demanding that he reflect, then the machineries of his brain will be compelled to fire up, and thereby work to do their (and by extension, Mr P’s) best to respond by releasing patterns of behaviour which (in time) work to convey what is going on re certain regularities and oddities in its (Mr P’s) world.

However, given that certain aspects of Mr P’s world are presenting as (very) abnormal, but language is through-and-through a medium of norms, so corolling in
linguistic boundaries the oddities which certainly seem to be going on in and around him mightn’t be a straightforward task. As with Dennett ordering beans, there will be an almost inevitable slide in the direction of over-specification. Moreover, once actualised, the over-specified patterns assume the role of pattern in noise in the world. Whereupon, in whole or in part, they may be selected-in as important data by Mr P’s information processing machineries, again and again and again. Thus progressively producing what has previously been described as a non-virtuous feedback circle. And on account of which, the oddities will be disposed to compound. So that eventually, if Mr P thinks back on his problems, he may well remember having seen or heard a second head with all the acuity of everyday perception. Accordingly, this is what he quite reasonably and honestly recounts to his physician as having taken place.

This is by now a familiar story which in its fundamentals hasn’t changed from Dennett and the Molly B, to Patsy, to Mme M, and to Mr A. Thus, it can at least be observed that a D-PE account seems on the face of it to submit a unified explanation of characteristically different kinds of case.

Yet at this point it might be challenged that I am being somewhat dishonest in so much as what I am suggesting as an answer isn’t really an answer, for the reason that it doesn’t explain Mr P’s hallucinations so much as deny them or explain them away as confusion or misremembering. However, in response, I must stress that the account neither denies nor explains away Mr P’s experiences. Consider that within the context of Dennettian theory, an interpretation of (very) strong hallucination in terms of its essentially being an overdeveloped narrative, i.e. ‘detailed fantasy’ or ‘self-perpetuating personal myth’, doesn’t necessarily mean that it is a story without ‘pictures’, i.e. associated experiences. Indeed, as was explained in chapter one, Dennett does not deny experience, only the conventional way of construing or defining experience as actual, unified, progressive, phenomenally conscious content which as a matter of fact is generated in toto via (supposed) dedicated experiential machineries in a living brain (Dennett 1991 [1993], 1992). Moreover, Mr P’s experiences are neither something that a D-PE account (at least as I present it) would want to deny, for a number of reasons, nor something that a D-PE account needs to deny, for the reason that it is well suited to explaining them.

So why are Mr P’s experiences not something that a D-PE account (at least as I present it) would want to deny? There really are a good number of reasons for saying this, but I offer three.
First, there is a reason from within the theory itself. Recall Dennett’s commitment to autophenomenology and heterophenomenology (e.g. Dennett 1991 [1993], 2002a, 2003, 2006a, 2007). Accordingly should Mr P maintain that he sees, hears, feels, etc. *this* second head; and if he is observed to interact with the second head in ways it would be reasonable for him to interact if the head were actually there; and if his behaviours re the head are reasonably consistent over time; and if there is no good reason to suppose that Mr P is the victim of trickery, or illusion, or that he is engaged in deceit; then in line with the precepts of heterophenomenology not only who are we to say otherwise, but to insist otherwise inserts a disjunction into the Dennettian thesis, because now we would need to differentiate between cases in which heterophenomenology is the standard, and cases in which heterophenomenology is unreliable and fails. And I suggest that no Dennettian would choose to do this.

Second, there is a reason supplied by clinical precedent. Recall DSM-V’s explicit instruction that:

*Hallucinations* are perception-like experiences that occur without an external stimulus. They are vivid and clear, with the full force and impact of normal perceptions, and not under voluntary control ... [and] must occur in the context of a clear sensorium ... (APA 2015 p. 87.)

And although Dennett is ‘particularly sceptical’ as to the veracity of reports of very strong hallucinations – which actually I am proposing that Mr P’s case of hallucination is, not least because it seems to me that matters can’t get much stronger than trying to shoot the antagonistic hallucination dead – he clearly observes that:

But if *really* strong hallucinations are not known to occur, there can be no doubt that convincing, multimodal hallucinations are *frequently experienced*. The hallucinations that are *well attested in the literature of clinical psychology* are often detailed fantasies far beyond the generative capacities of current technology. (Some italics added, Dennett 1991 [1993] p. 7.)

Third, there is a reason supplied by the fact of being with patients at the time they are hallucinating. That is to say, when I worked as a nurse, on numerous occasions I was with a patient while he or she was hallucinating ‘whatever’. It happened across a range
of circumstances, for instance, in cases involving delirium, drugs, psychosis, dementia and death. And an outcome of this is that, I have little doubt that, whatever else is going on in cases where people are hallucinating ‘whatever’, all things being equal, they are experiencing something, and in some cases at least they certainly seem to be experiencing that something in a way which is at least compatible to the way in which they are experiencing other normal contents at the same time. For instance, consider the farmer who in his small hospital room was concerned for my welfare lest the cows which were milling around us would bump into me. Moreover being with and talking to people in these situations, in many cases it is evident that a patient’s responses re his or her hallucinatory experiences exhibit a coherence and rationality re the content of the hallucination, not only like the farmer at a particular moment, but in some cases extending across (quite lengthy) periods of time. For instance, take the cases mentioned in chapter two, of Ethel, and her companionable hallucinatory voices, and Andrew, who maintains that he can see through space and time, and at that point had a lengthy medical history of insisting that this was so. Therefore, from the point of view of a D-PE account, if Mr P says that he sees, hears, feels etc. a second head, then there is no question of denying or explaining away that from his point of view he is experiencing something of note.

Finally, why is it that Mr P’s experiences are not something that a D-PE account needs to deny, for the reason that it is well suited to explaining them? To explain, first consider what the account says in the normal case.

The D-PE model rests on the idea of a motivated system sharing ever-changing informational patterns across itself by way of top-down concepts (expectations, predictions) functioning to select-in patterns from bottom-up noise (error) countless times over across a system which at every level is deriving conclusion upon conclusion as it strives to negotiate its world.

Somewhere between its bottom and top stages of processing the system derives a class of conclusions which are special because they are instances of content fixation which in the right circumstances release patterns of downstream changes which work to inform the system (potentially) at the level of the personality of certain (likely) criteria in the surrounding (observable) world, whereupon the subject may see the tree, hear the birdsong, feel the breeze.

The subject seeing, hearing etc. happens not via real changes in actual consciousness in virtue of which the subject sees etc. x, but via a series of behavioural
adaptations which work to inform the system at a certain level of function and via a certain mode of transmission about specific (likely, important) patterns in its surroundings, i.e. perceptual field. Accordingly, there is nothing-it-is-like for one to see, hear etc. That is to say, there is no actual conscious content in virtue of which a living subject experiences whatever. Instead there is informational content, updating the system in a mode which is fast reliable-enough and amenable-enough to the system at the level of personality, and in virtue of which from a first-person perspective there is something-it-seems-like to witness the properties of the world.

At no point is there a polished final product of experiential content. The pertinent content fixations in the brain’s own story are a raggedy hotchpotch of shifting inferences at various stages of drafting and scattered across different locations. This can be compared to the litter of content re James Joyce’s room, from *Finnegans Wake*. Nevertheless, in the push to make sense of what is happening, i.e. derive a coherent enough fiction to be useful, the system – and eventually perhaps the subject – submits these raggedy impressions to further rounds of higher-order redrafting, which (redrafting) potentially imposes coherent enough edges to potentially allow the system/subject to focus on a particular pattern or patterns through the noise. A lot of top-down expectation/prediction goes in to this task. (Cf. Dennett’s account of the eye-tracker experiment, Dennett 1991 [1993] pp. 361-2.) And the outcome is something like an ongoing experiential story of many shifting and competing parts. That is to say, the meandering ‘variegated jumble of images, decisions, hunches, reminders and so forth’ reminiscent of the first-person stream of consciousness depicted in another of Joyce’s novels, *Ulysses* (Dennett 1991 [1993] pp. 212-4 etc).

Now consider the case of hallucination. By definition, it is a first-person (subjective) perception-type experience in the absence of a significant third-person (objective) stimulus. Yet within the D-PE framework, even in a normal case of perception, given the way in which once selected-in patterns are repeatedly broken and reshaped via the system’s highly competitive inferential machineries, so there is no necessary connection between a subjective experience of $x$ and the actual presence of a related $x$-object or $x$-property in world. However, in the case of hallucination, there is noise. And specifically, not just any noise – because even at the best of times this system is permeated by noise – but there is slippery noise which masquerades as viable pattern. That is to say, if in order to meet its requirements in special or difficult circumstances a system has opened itself to epistemic risk, then it might be the case that, driven by need,
and reasonably enough, the system errs by deriving meaning where truly it ought not. Thus without something in place to check or constrain the error, the noisy contents of the duplicitous pattern may thrive and gain purchase at pertinent levels. So that potentially, given the ‘right’ circumstances, it is feasible that erroneous contents could acquire a place in the brain’s narrative and so, subsequently, start playing out.

Add to this the predisposition of a human system for words. That is to say, all things being equal, it is likely that to start with the hallucination-type experiences will be a raggedy hotchpotch of shifting impressions at various stages of drafting. Maybe something and nothing, which in normal, circumstances would be ignored. However, in certain circumstances, such as delirium, agitation and/or confusion, they are not ignored. They are taken as given. Whereupon, driven by the need to make sense of his or her situation, the subject could well be moved to impose answers on them, round after round. Yet given the oddness of the content, so the norms of language inevitably will be stretched. So that in conjunction with a push towards over-specification, the conclusions which are drawn will likely get ever more unreal.

Yet, in any case, given the motivation and content involved, so the experiences and conclusions are likely to feed back into the system as meaningful information time and time again. Each cycle adding verisimilitude to an increasingly detailed stream of conceptions – at various levels – and so which content will likely be getting stronger, i.e. more and more ‘objective’ or seemingly real (Dennett 1991 [1993] p. 7-13).

So is all of this sufficient so that Mr P sees a second head? The answer is, Yes. And in theory it can happen in several increasingly ‘strong’ ways.

In the weakest sense it might be that Mr P experiences something nearer to an intuition that there is a head, which would be sufficient for him to know deeply that there is a head, but which lacks the colours, textures etc. of his typical experiences of the everyday world.

Then in a middling sense – and likely after a period of head-related intuition-like contents feeding back in as ongoing evidences and so supporting the story – if content fixations supplying the patterns for experience are invested with sufficient energy and information, then should the brain be informed that there is a head, so, if other conditions hold, Mr P will see a head, and plausibly ‘with the full force and impact of normal perceptions’ (APA 2015 p. 87).

Finally, in the strongest sense, if perceptions of the head along with all other evidences continue feeding head-related patterns with energy and information, not only
could Mr P see the second head, but he could experience it as being lit up by salience across different modalities. Whereupon, if Jesus and Abraham were going to talk about anything, then it would be about the head, wouldn’t it? Really, what else might there be to talk about? In other words, across Mr P’s system, head-related content is now foregrounded to the nth, so quite plausibly his corresponding hallucinations are very strong indeed.

At no point has this account of Mr P’s hallucinations strayed outside the D-PE model. Yet specifically re Mr P’s case, consider that his is a system which has been under a great deal of stress for a substantial period of time. At the very least, we know that Mr P’s personal circumstances are tragic; that he has a long history of psychiatric disorder; and as regards the head, Mr P (his system) has been progressively accruing evidences across different modalities – including the corroborations supplied by Jesus and Abraham – in addition to which are those evidences supplied by his own (dodgy or not) memories of events. So that, as incredible as Mr P’s case might seem to you and me, the D-PE model seamlessly tolerates Mr P’s assertion that he can see etc. a second head.

A further point to note is that, even if it is the case that Mr P is experiencing the very strongest hallucinations, the D-PE model would maintain that Mr P’s system is not operating at random. That is it say, it is not mere chaos inside him. This is because throughout the D-PE system, information processing is held to be organised via an internal logic which according to Dennett is ‘not apparently algorithmic, but rather at best heuristic’ (Dennett 1978 [1981] p. 303), and which in this thesis I have specified as being consistent with the rules of prediction error reporting. Moreover, all the while, Mr P’s system is continuing to process all other non-head-related stuff precisely as it ought to. And this point has two consequences. One, it explains why Mr P’s head stuff occurs within the context of a clear, i.e. otherwise normal-enough, sensorium (APA 2013 pp. 87-8). That is, how and why the head-related narrative dovetails with the normal one(s). Two, it explains why the hallucination exhibits coherence and cohesiveness across time. That is to say, it is sustained by a deep internal logic which is fundamentally the same as the deep internal logic sustaining all of Mr P’s other thoughts and experiences, and so for as long as the head-story continues being supplied with energy and (corroborating) information, so the D-PE model predicts that it should continue playing out through time.
Finally, the closing question has to be, after he shot it and it was hanging by a thread, why did the head cease to bother Mr P? In response, all I can do is speculate that: an ‘organic’ answer could be that, in shooting himself in the (his actual) head, Mr P could have damaged part of the neural machineries on which his brain’s own head story depended; a ‘prediction-error’ answer might be that, when you shoot something in the head there is a deep expectation that it will ‘go away’; and a ‘psychodynamic’ answer might be that, the fact of shooting himself in the head was sufficient to sober up Mr P, regarding the head story at least.

Concluding Considerations

I start this thesis by thinking about the problem of delusion, and in so doing identify two questions to be addressed. The first question asks, which theoretical approach to delusion most comprehensively addresses the data set of features of clinical delusions. Following on from this, the second question asks, is the account most capable of addressing the data set of clinical features an empiricist approach, a rationalist approach, or something else. To answer these two questions, I select three cases of delusion which together are representative of the wider data set. The delusions I consider are: erotomanic delusion, Capgras delusion and perceptual delusional bicephaly.

Then in chapter two I analyse the three prevailing theories of delusion. These are: the empiricist one-factor account; the empiricist two-factor account; and a relationist rationalist approach to delusion. I examine each theory both in itself and against the test cases of delusion. The conclusion I draw is that as regards the rigid stipulations which each theory respectively adheres to concerning a set of foundational criteria, all of the prevailing approaches to delusion run into an intractable knot of difficulties when it comes to addressing the test group of delusions.

Supported by this conclusion I turn to consider an alternative approach. The approach combines Daniel Dennett’s theory of mind with the principles of Prediction Error theory in the brain. The outcome of merging the two discrete theories into a single explanatory framework is that it delivers a wide deep rich explanatory model concerning how and why a motivated system continually acts within itself and its surroundings to reduce the mass of information potentially available to it as by necessity it negotiates the changing circumstances of its world. I suggest that by taking a normal ‘control’ D-PE model, and interleaving it with the various ‘abnormal’ test cases of
delusion, one might be able to specify the points within the system where error sets in, and moreover, predict and track how the inconsistencies take hold.

Finally, applying the D-PE model to the test cases of delusion, I reason that the approach comprehensively addresses the key features of each of the test cases, which it systematically accounts for in terms of the special or difficult circumstances which the subject of the case is navigating in conjunction with a central role being awarded to the introduction into the system of noise. Thus re the first thesis question, my answer is that a D-PE approach most comprehensively addresses the data set of clinical features of delusion. As for the second question, my answer is that, although technically speaking a D-PE account affords a rationalist approach to delusion, it does not seem correct to pigeonhole it as a rationalist account as rationalism re delusion is typically construed. This is because, pace the standard (relationist) rationalist model, a D-PE approach holds the subject to be a deeply rational human being, as well as it wholly allowing that in any particular case abnormal experiences could have a key explanatory role. Thus I suggest a more realistic answer to the second thesis question is that a D-PE approach to delusion occupies a space between the conventional empiricist and rationalist models, while standing out as substantively different to either. Therefore overall I submit that a D-PE approach provides a legitimately alternative way of thinking about the problem of delusion and which given its explanatory potential is deserving of consideration alongside the more established views.

To close, it might be asked whether a D-PE approach is applicable only to cases of delusion or whether it might have wider application in the clinical sphere? My answer is to strongly affirm that as the approach hinges on the idea of a deeply normative Dennettian system, i.e. a living human being who by necessity needs to navigate the facts of his or her world, and which navigations can at any time go ‘right’ or can go ‘wrong’, so the approach seems ideally placed to talk about all manner of clinical concerns. Moreover, the inclusion of prediction error reporting into the Dennettian framework provides a way of talking about pathways by means of which irregularities can enter, travel or compound. As regards specific areas of clinical interest, while writing this thesis, certain fields have stood out to me as being particularly viable areas for possible further research.

The first area relates to the effects on the system of drugs. For instance, Phil Corlett, Chris Frith and Paul Fletcher (Corlett et al. 2009) talk about how the principles of Bayesian prediction error reporting in the brain can further understanding of the
symptomatic effects that psychotomimetic drugs (e.g. ketamine and LSD) have in healthy volunteers. However, I submit that a way of furthering this idea would be to think about the problem of drug use from the perspective of a wide deep rich D-PE system, re actual users, and patterns surrounding specific drugs and drug use, not only with regard to an individual subject’s biology, but also in respect of his or her reasons and behaviours, both personally and re the multiple environments with which the person interacts.

The second area relates to the changes and challenges faced by a human system as it is pushed far beyond its normal tolerances around or at the time of death. As regards research in this area, I believe there is no better place to start than with anything written by Elisabeth Kübler-Ross (e.g. Kübler-Ross 1969; Kübler-Ross and Kessler 2005). Yet also of significance from the field of near death studies is research into Near Death Experiences (NDEs), which on a personal note I take to be especially meaningful not least for the reason that in 2001 during surgery I underwent a NDE. I suggest that a potentially fruitful starting point for approaching this issue from a D-PE standpoint would be the research by Kenneth Ring and Sharon Cooper into NDEs in the blind (Ring and Cooper 2008).

Beyond this all I would like to add is: Thank you to all of my case studies from this chapter, but especially thank you and good luck Mr P!
REFERENCES

Boon, J. and Sheridan, L. editors (2002) Stalking and Psychosexual Obsession: Psychological Perspectives for Prevention, Policing and Treatment, Chichester: John Wiley and Sons Ltd.


Hughes, T. (1989) Moortown Diary, London, Faber and Faber. (Originally published as Moortown by Faber and Faber, 1979.)


