The nature of delusion:
An analysis of the contemporary philosophical debates

by

Aline Aurora Maya Paredes

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy at the University of Central Lancashire

June 2017

STUDENT DECLARATION

Concurrent registration for two or more academic awards

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution.

Material submitted for another award

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work.

Signature: Aline Aurora Maya Paredes 

Date: 09/06/17

ABSTRACT

The present thesis surveys different philosophical approaches to the nature of delusions: specifically, their ontology. All of these accounts face general problems; my aim here is first to illustrate them in the main body of the thesis and then, in Chapter VI, to show why this is the case. First in this work I examine Jaspers' two accounts of delusions; one regarding their characteristics and the other one portraying delusions as the ununderstandable. This sets the challenge for the subsequent philosophical accounts of the nature of delusion which aim to shed light and some understanding – contra Jaspers – by giving a more substantial account of what sort of mental state delusions are. I challenge this conception by analysing the underlying assumptions of the most representative accounts of delusion.

In this work I consider characteristics of antidoxastic, doxastic, and phenomenological accounts; I also devote a chapter to the conditions of possibility of propositional attitudes. However, since none of the various theories of the nature of delusions succeeds, I argue that there must be something problematic about the form of the analyses commonly offered. My argument proceeds by comparing the latter to Pickering's critique of the 'likeness argument'. Approaches of the form of the likeness argument fall prey to an intrinsic question-begging that undermines the very same account. The general problem seems to be that all the stances attempt to approach the debate by comparing delusions to more familiar kinds of mental states; this, faces the objection of making delusions too easily understandable. On the other hand, retreating from this by defining more bespoke, less familiar mental states threatens to make the accounts ad hoc and unilluminating. My general conclusion is that one cannot characterize delusions without taking away what it is distinctive about them.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I have been able to engage in this endeavour due to the scholarships granted by the Consejo Nacional de Ciencia y Tecnologia (National Council of Science and Technology) in Mexico and by the University of Central Lancashire in the United Kingdom.

I sincerely appreciate the valuable help and comments from my two supervisors, Prof. Tim Thornton and Dr. Gloria Ayob.

I would like to thank the support in all aspects from my parents, Agustin Maya, and Ruby Paredes.

Special thanks to Annette, Sandra Rouse, and Dr. Mark Speakman.

INTRODUCTION

The aim of this introduction is to show the relationship of the present work with the contemporary literature of delusions. Likewise, it states the main objective of this thesis and gives a brief summary of each of its chapters.

Delusions are amongst the most puzzling mental phenomena. What they are, what they mean or perhaps what they are trying to express is something deeply concerning, that nobody seems to be able to interpret literally. They, however, appear to be sincerely held, and extremely important to the person who is suffering the delusion.

It is through the diagnosis of having a delusion that an individual can be classified as psychotic, but it has been impossible to pinpoint what is the specific feature that is believed to be wrong in that person's mind. Since its classification as a symptom, there has been much debate about the nature or the ontology of delusions because they have evaded all sorts of characterization; and this issue remains pressing despite advances in psychiatry and the neurosciences.

Philosophy, and in particular the philosophy of delusions, is an important resource for psychiatry, as Karl Jaspers remarked:

> Many a psychiatrist has said that he did not want to burden himself with a philosophy and that this science had nothing to do with philosophy…But the exclusion of philosophy would nevertheless be disastrous for psychiatry… If anyone thinks he can exclude philosophy and leave it aside as useless he will eventually be defeated by it in some obscure form or other…Only he who knows and is in possession of his facts can keep science pure and at the same time in touch with individual human life which finds its expression in philosophy. (Jaspers, 1913, pp.769-770)

Jaspers' views greatly contributed to the way in which the particular definition of delusions can be found in the Diagnostic and Statistical Manual of Mental Disorders (DSM) and elsewhere. Berrios (1991) observes that some notions about delusion previous to Jaspers can be found in the 17th century in the work of Thomas Hobbes and John Locke.

However, a new, fresh philosophical discussion of delusions emerged with a different approach since Maher (1974). This debate had a boom through the 1990's and there has been further sophisticated discussion through the 21st century. But, the main question remains: what are delusions? It is obvious that a standard definition of delusions exists, but this definition is far from satisfactory.

Through this introduction I outline some key examples of delusion and some of the dominant frameworks that will set the scene for the discussion in the various chapters of this thesis concerning the nature of delusions.

I. Definition of delusion

Delusions are usually the result of a brain injury, especially to the right-hand hemisphere of the brain[1], or a symptom of a neuro-degenerative disease such as progressive dementia (Breen, Coltheart, et al, 2000). Delusions can be, more specifically, seen in schizophrenia, paranoid schizophrenia, schizoaffective disorder, affective disorder, Alzheimer's disease, Lewy body dementia, multi-infarct dementia, head trauma, epilepsy, cerebrovascular disease, pituitary tumor, multiple myeloma, multiple sclerosis, viral encephalitis, frontal lobe pathology, and AIDS (Coltheart et al, 2007).

They have been defined by the American Psychiatric Association at the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), as follows:

> Delusions are fixed beliefs that are not amenable to change in light of conflicting evidence. Their content may include a variety of themes (e.g., persecutory, referential, somatic, religious, grandiose)[...].
>
> Delusions are deemed bizarre if they are clearly implausible and not understandable to same-culture peers and do not derive from ordinary life experiences. […] The distinction between a delusion and a strongly held idea is sometimes difficult to make and depends in part on the degree of conviction with which the belief is held despite clear or reasonable contradictory evidence regarding its veracity. (American Psychiatric Association, 2013, p.87)

[1] According to Breen, Coltheart, et al. (2000), although there are many different conditions associated to delusion, there seems to be a strong correlation to some delusional syndromes and right-side hemisphere brain injury.

This standard definition is the one currently used to diagnose psychotic disorders in patients. The definition starts by pointing out that delusions are *beliefs*. This aspect is nowadays the centre of much debate and will constitute the core of two chapters in this thesis. It is also not satisfactory to say that delusions are not amenable to change, since some patients have been able to recover from their delusion. As the current DSM 5 definition points out, delusions do not need to be bizarre, although often they are. And as the definition also states, to distinguish a delusion from a strongly held idea is not straightforward. So, in total, we are left with very little to actually define what a delusion is. But that does not mean that no effort has been made to fix the problem; on the contrary, there have been several proposals, and many have been important or influential, so much, that it is hard to say which one is right. Therefore, in order to not lose clarity about this matter, it is worthwhile to look at real examples of delusions.

II. Examples and cases of delusions

It is useful to divide delusions into two rough categories: schizophrenic delusions, which tend to be florid, elaborate, and are usually composed of various related hypothesis; and monothematic delusions, where the delusional thought is relatively constrained and involves a single theme, whilst the rest of the subject's mental life remains relatively spared. That means they are usually composed of a single delusional idea that has a main theme. Usually, monothematic delusions are caused by some sort of organic brain damage, whilst schizophrenic delusions are caused by schizophrenia. It should be noted that neither of these two categories is definite, because schizophrenic delusions might be constrained to a single theme whilst monothematic delusions can sometimes have a variety of related ideas. However, the division has been useful to understand delusions better in philosophical discussions. As examples of monothematic delusions there are the following:

a) Monothematic delusions:

Capgras delusion

Discovered initially by Joseph Capgras, a French psychiatrist, in 1923, this delusion centres on the patient's idea that one of her close relatives, usually the husband or the wife, has been replaced by an impostor. It is important to note that the figure replaced is someone very close to the patient. The supposed impostor could be thought to be an alien, a clone, a robot, an actor, or anyone pretending to look like the supposedly replaced relative. Patients usually say that they can perceive small differences that nobody else would spot, such as the supposed stranger being smaller, thinner, taller, or having a different way of talking or acting. This delusion causes much distress to the patient, and she might sometimes engage in violent behaviour towards the stranger. One study showed that the latter happened only in the minority of the cases, approximately in 18% of a total of 260 cases analysed (Breen, Coltheart et al., 2000, p.80). In most of the other cases, the patient acts in a friendly way to the supposed impostor. The first extensive recognized account of Capgras was published in 1923 (Capgras and Reboul-Lachaux, 1923).

Ellis and Young (1990) suggested that the Capgras syndrome was a mirror disease of prosopagnosia. Prosopagnosia is a condition in which the person is unable to recognize people, even if they have known them for many years. In prosopagnosia, the affective response is unimpaired, and people are able to feel affection for their loved ones even when they are still unable to recognize them and cannot tell who the person they are looking at is. But it was noted that Capgras patients do not suffer from prosopagnosia: "prosopagnosic patients do not usually experience the Capgras delusion and, conversely, Capgras patients can still recognize a number of familiar faces" (Young et al., 1993, p.697). Patients with Capgras are able to recognize the supposed impostor as identical to the person who was allegedly replaced. Ellis and Young's (1990) theory implies one of the most important findings in Capgras: patients would show a reduced autonomic response to faces when their skin conductance is measured; this theoretical implication was confirmed by experiment (Ellis et al, 1997). In practical terms, this means that, although they are able to recognize the person as someone with the same characteristics as their loved one, patients with Capgras will not feel an appropriate affective response towards them (Ellis and Young, 1990). This, according to the theory,

is the consequence of there being two different independent routes in the brain for face processing, one of them which could be damaged in prosopagnosia, and the other one in Capgras. These routes had already been proposed in previous works, but Ellis and Young applied it to Capgras patients: "This work by Ellis and Young was based in part on a two route theory of face processing proposed initially by Bauer (1984, 1986), in which both a visual route (…) and an affective route (…) contribute to the face recognition." (Breen, Coltheart et al. 2000, p.77) Due to the fact that the affected route in the brain is a visual one, it was theorized and predicted by Ramachandran (1998) that some patients with Capgras would be able to interact normally with their loved ones when they spoke to them by telephone, and this was experientially confirmed.

Patients with Capgras usually realize that their claim about their loved ones being replaced by an impostor is very bizarre, although they still keep having the same fixed thought. Young explains this in his (1998) book:

> if you ask "what would you think if I told you my wife had been replaced by an impostor" [to a delusional patient] you will often get answers to the effect that it would be unbelievable, absurd, an indication that you had gone mad. Yet these patients will claim that, none the less, that is exactly what has happened to their own relative. (Young, 1998, p.37)

The following is a clinical case study from a patient, called Fred by the researcher, suffering from Capgras (Lucchelli et al, 2007):

> Fred's wife reported that about 15 months from onset he began to see her as a "double" (her words). The first episode occurred one day when, after coming home, Fred asked her where Wilma (a fictitious name) was. On her surprised answer that she was right there, he firmly denied that she was his wife Wilma, whom he "knew very well as his sons' mother", and went on plainly commenting that Wilma had probably gone out and would come back later. Fred's wife was puzzled by her husband's behaviour, but when, some 15 minutes later, he was normal again, she thought it must have only been a joke. She started to worry when over the following weeks similar episodes occurred with increasing frequency. At the time of the first visit they were occurring almost every day and

> even several times a day. As she reported, without any apparent triggering event, Fred started to behave toward her as if she were not the real Wilma, but another woman who, as he himself stated, bore the same name, had the same physical appearance and the same voice, wore the same clothes, lived in the same house, but nevertheless was a different Wilma. In some cases the disappearance of his "real" wife upset him: he repeatedly asked where she was, obviously worried about her. On one occasion, despite his wife's strong protests, he left the house and went looking for her in the streets; when he came back after a few hours, he looked relieved to find her at home and anxiously asked where she had been and why she had not told him she was going out.
>
> On another occasion, he urged her to go with him to the police to report Wilma's disappearance. Most of the times, however, he seemed quite pleased to see her as the "double" Wilma and addressed her in a very gentle way. His wife described his manner as "courting as when we were dating". (…) For several months he had no difficulty in recognising other people (…)(p.189).

It has been found that subjects with Capgras still have intact most of their reasoning capacities. Tests for logical reasoning have been carried out with deluded patients and they compare well with normal controls. In patients with Capgras their capacity for giving reasons, other than the delusion, seems relatively spared.

The Cotard delusion

In the Cotard delusion, subjects express the conviction that they are dead or non-existent; they might say that they are immortal because they are already dead, or that their internal organs are rotting or missing. Patients do not mean this in a figurative way; what they are trying to communicate is that they literally believe they have died, usually from some minor disease they have had in the past. Subjects with Cotard are always found to suffer from deep organic depression (Yarnada et al., 1999). The original report of the Cotard syndrome was made by Jules Cotard in (1880) where he presented the case of a 43-year-old woman who claimed that she had no entrails, that she did not have need for food, that she was just skin and bone, and that she was eternal. This patient

also said that she needed to be burned alive, and had made several suicide attempts (Berrios and Luque, 1995).

Some studies have found a case in which a patient suffered from Cotard and then it turned into the Capgras delusion when her mood changed (Wright et al., 1993). Young (1999) reports:

> This association of two unusual delusions has been reported in other cases too, and the key factor seems to be the patient's mood – when in a suspicious mood they seem to think that other people are impostors, when depressed they think they are dead. There is an obvious parallel here to findings that people with persecutory delusions tend to make external attributions and depressed people internal attributions as to the causes of negative events. (Young, 1999, p.577)

Young (2000) claims that a clue to understand this phenomenon comes from studies that have shown that depressed people tend to attribute negative experiences to themselves, whilst people with persecutory delusions attribute negative events to external causes (Candido and Romney, 1990).

One characteristic case of Cotard is reported in (Vaxevanis, 2005):

> Patient B A; This is a 46 years old female patient, the youngest of six brothers and sisters in her family. She arrived in the Emergency Department Unit of the "Psychiatric Hospital of Thessaloniki" in 1999, escorted by two of her brothers, who reported having found her naked on the balcony of her house ready to jump off, and that lately she was melancholic and very afraid. During her treatment, the patient complained that nobody could help her because she claimed she was a "dead – plant", thus neither alive nor dead and that she would remain in this condition eternally. Characteristically, during her interviews the patient reported:
>
> "I have not eaten for months / or gone to the toilet/ all the organs within me have rotten/ the food can't pass through, everything has been coagulated /the Lonarid (paracetamol) that I took, have stuffed my bowel, this is unfortunately my punishment / I am tired, I haven't slept for years/ I have no blood/ I have no heart, it doesn't beat anymore. I was deceived at the ECG department while they

> knew that my heart doesn't beat anymore/ you are deceiving me when you take my blood pressure, because I'm not alive anymore, I'm a dead-plant". (Vaxevanis, 2005, pp. 41-42)

Fregoli delusion

Patients suffering from Fregoli express the claim that someone they know follows them all the time, but that this person is unrecognizable to others because she is in disguise. Ramachandran and Blakeslee (1998) propose that it is plausible that subjects with Fregoli may have a heightened affective response to unknown faces. One case of Fregoli with several misidentifications of strangers as known people is described in Feinberg (1999):

> BJ reported that a man he used to work with was in the hospital as a patient in a wheelchair. He believed that this co-worker was calling himself by a different name, which was actually the real name of another patient hospitalized on the unit at the same time. BJ described how,"I busted his chops about it a few times, and then my wife was talking to him, and she told me leave the guy alone: 'He's got the name. Don't cause him any problems here.' And I agreed with her. It's not the thing to do. I get out and he gets out, he's gonna wind up back with me [at work] and that's it". (Feinberg, 1999, p.379)

BJ was able to recognize the physical differences as well as similarities between his co-worker and the actual patient who was misidentified, on one occasion stating, "He used to have light blond hair and he always had a big beard". Yet, he insisted that the two shared the same identity.

A second, work-related, misidentification occurred when BJ reported that another patient in the hospital was a former female client. Again, she had claimed that her name had been changed because he was aware that other people referred to her by another (i.e. actual) name. He stated that

> There's a woman here who was uh, was my customer contact… I've known her for twenty years. A very nice woman, and when I tried to talk to her… the people here got upset and told me you're tryin' to flirt with the girl." He added "They

were sayin' that I was flirtin' with her, and I wasn't 'cause she was here with her husband. (p. 379-380)

Erotomania

In erotomania, or de Clerambault syndrome, the subject believes someone, especially a famous person or someone from a higher social status, has fallen in love with him or her. In this delusion the subject maintains that this love is demonstrated by encoded secret messages that only the patient is able to decipher. The patient usually believes that it is the other person who was the first to fall in love, and usually tries to correspond with this supposed romantic partner with letters, gifts, phone calls, messages, etc. It was first described in its modern form by de Clerambault in (1942). A representative case of a patient with erotomania is described in (Kovacs, 2005):

A 34-year-old, single, male teacher was admitted for the first time one week after his suicide attempt. (…) The object of his delusions was a married woman from a well-off family, whom he had met while studying at university. He had made an attempt to start a relationship but was rejected, and a relationship never evolved between them. (…) He broke all contact with his family after finishing university, and he emphasized this decision by changing his last name. Subsequent to these events his erotomanic symptoms evolved right after he had met some people known both to him and the woman. He felt that the woman fell in love with him, and she tried to let him know this through these people, in a hidden, but to him, quite evident way. In retrospect he re-evaluated their rare, earlier contacts, influenced by his erotomania. The patient kept in touch with her by writing letters and calling her on the phone. The woman and her family did not respond to his phone calls or letters in any way from the beginning, but he discovered secret messages and instructions sent to him by the woman, through certain signs and voices of the telephone system (e.g., ''engaged,'' ''at the moment not available''). He then tried to make personal contact with her by turning up at her home, and he started to harass her with marriage proposals. These efforts were met with increasingly harsh refusals. In spite of these responses the patient was convinced that the woman and her parents and sisters expected him to marry her. (p.370)

Anosognosia

In anosognosia patients have the thought that they are still able to move a paralyzed limb or part of the body, and that they can carry on doing their duties normally, as if this paralysis did not exist. This condition is usually found in patients who have suffered from severe sudden brain damage, for example, a stroke. The condition is only persistent in patients with damage to the right-side of the brain, although people with some left-side damage can present it too for a short period of time. A clinical report of anosognosia can be found in (Venneri and Shanks, 2004):

> This is a case report of an 85-year-old right-handed Scottish woman, EN, who showed persistent anosognosia for hemiplegia following a haemorrhagic stroke. Extensive damage in the right hemisphere caused left upper and lower limb flaccid hemiplegia and severe left-sided neglect. Lack of awareness for her deficits was still present 2 years after the stroke, when neurological, neuropsychological, and SPECT examinations were performed. Testing revealed severe left unilateral neglect and poor performance on verbal fluency tasks. EN had age normal memory performance, and her object recognition and praxic abilities were preserved. She showed no global reasoning or language problems apart from her abnormal beliefs. EN believed that she was able to walk and carry out several activities, in a context of other disorders of belief. SPECT scan showed marked hypoperfusion in the right parietotemporal cortex and this extended to the associative cortex in the right frontal regions. The persistence of anosognosia in this patient cannot be explained by memory impairments or global cognitive decline. (Venneri and Shanks, 2003, p.230)

This delusion is philosophically interesting due to the fact that only certain people, that is, people with damage to the right hemisphere, get the delusion. This demonstrates that the delusion is not due to a general malfunction of the brain; it is a very specific injury, which explains why other cognitive abilities, such as memory and reasoning, seem to be intact.

b) Schizophrenic delusions

Delusions that result from schizophrenia are usually florid and cannot be classified as having a single delusional content, as usually happens with the examples of delusions mentioned above. However, there are some common delusional thoughts that are recurrent in schizophrenia, such as:

Paranoid delusions- In these delusions the subject believes that total strangers, a secret government agency, doctors and nurses or the police, is trying to harm her. Case report from (Basu, 2013):

> Ms. B, a 26-year-old woman, presented to our clinic with insidious onset psychiatric illness of about five years duration. She was educated until intermediate level and belonged to a low socioeconomic status Hindu joint family of urban background. She had no past or family history of psychiatric illness, and she had good premorbid functioning. Her symptoms started with academic decline, suspiciousness, and referential ideation progressing to gradual social withdrawal, predominant irritable mood, and delusion of persecution by her college authorities and employer. A year after these initial symptoms appeared, she began to exhibit odd behaviour such as sewing clothes in strange patterns, offering worship by performing strange rituals, and riding a tireless bicycle at midnight for no apparent reason. She had disturbing auditory hallucinations that were commanding type and voices commenting on her actions. Auditory hallucinations were followed by visual hallucinations in which she would see a man and woman entering inside her body and controlling movements against her will. (…)During the third year of her illness, she developed persecutory delusions against family members—that they were conspiring to take her property. She alleged they poisoned her, causing her to lose her voice. Hence, for the last three years, she completely stopped speaking and would communicate everything by gesturing or by writing. Thereafter, she deteriorated in that she was irritable, had impaired selfcare, slept very little, and ate minimal amounts of food because of fear of harm by family members. About six months before admission she lodged a complaint with the police and filed a court-case against them. The legal services asked her to get a psychiatric

> assessment performed, which led her to approach our psychiatric outpatient services where she was promptly admitted. There is no history of jaundice, organicity, substance use, or sexual exposure. (Basu et al., 2013, p.11)

Reference delusions- In these delusions particular objects get a hidden significance that only the patient can decipher or interpret. Case report from (Bloy, 2011):

> "Brian" was a 32-year-old Caucasian British male who was referred to early intervention psychosis (EIP) services following concerns about his increasingly suspicious behavior. Brian lived by himself; however, he spent a considerable amount of time at his partner's home, who lived nearby. At the time of the start of therapy, he was enrolled on a full-time horticultural course that he had been undertaking for the past 12 months. (…)
>
> During the initial assessment, Brian described a range of psychotic symptoms, including paranoid delusions, delusions of reference, thought broadcasting, and thought insertion. (…) He also recalled having experienced ideas of reference, believing that television advertisements and billboards were relaying unique messages to him. Similarly, song lyrics held significant personal meaning. He became increasingly preoccupied with others' behaviors and words, interpreting them as evidence for his paranoid beliefs. He would spend a considerable amount of time ruminating about these thoughts such that this would impede on performing daily activities. As such, he found it increasingly difficult to engage in social situations, and he described frequent attempts to plan his daily schedule to avoid group situations or crowded places that may trigger paranoid and anxious thoughts. In the intervening years, these symptoms fluctuated in severity, at times abating while at other times occurring with a delusional intensity. (…) Social anxiety was prominent and avoidance arose from fears that strangers might approach him and that he might "freak out." Brian frequently structured his daily schedule so that he would minimize contact with others to avoid these experiences. He found it difficult to engage in daily tasks such as shopping due to the risk of contact with others. (Bloy, 2011, p.349-350)

Grandiose delusions- The patient usually states that he or she is a person of great importance and that he or she might have the solution for a very important problem, without having any evidence for it. Case report from (Lai, 2013):

> [Note: GD stands for Grandiose Delusions]
>
> A 30-year-old woman was referred to the psychiatric service after experiencing an uncontrollable body tremor in March 2010. About 3 months before the referral, the patient's unstable emotional and mental state was noticed by her mother when the patient experienced uncontrollable somatic discomfort. She associated these unusual experiences with the idea of being possessed by a devil. The patient described herself as an introverted and timid person with low self-confidence during her childhood and adolescence. (…) One day, the patient saw some flashing images, which she thought related to Jesus. She inferred that she was 'Jesus' after attending some augury sessions. She developed a bizarre delusion that she had a special ability to talk to animals and to predict the future. She also heard voices intermittently telling her that she could save the world. Her mother reported no significant mood changes to suggest mood swings or manic episodes such as pressure of speech or unusual increases in energy level. The bizarre delusion was maintained for about 2 years until her place of work closed. After starting another job, she thought that she had had bad luck and felt that her special ability was diminishing. She was criticised and her job was terminated by another company because of her unsatisfying work performance. Negative emotions developed and her self-esteem was fluctuating. One year before treatment, she was baptised and she destroyed the augury props. She started to experience uncontrollable somatic discomfort and saw black shadows moving around her almost every day. The content of the voices changed, telling her that she was being possessed. A delusion of demonic possession developed. The duration of untreated psychosis (the duration between the first appearance of the GDs and her first psychiatric treatment) lasted for about 3 years. (Lai et al., 2013, p.161)

Somatic delusion- The patient is preoccupied with having a very strange disease, and her claims for saying that are clearly bizarre. Case report from (Kepska, 2011):

> A 59-year-old woman was admitted to our dermatological ward because she was convinced that there were multiple purulent skin lesions on her back that had spread to her internal organs. She maintained that these symptoms first appeared 7 years previously, following her admission to hospital after a mosquito bite. The patient complained of numerous disorders during the previous years for which she had received treatment from various specialists (e.g., endocrinologists, gastroenterologists, gynaecologists, urologists) without seeing any improvement. A general physical examination and laboratory tests did not reveal any abnormalities. Dermatological examinations did not demonstrate any significant skin lesions besides a small scar on her back, which formed as a consequence of intensive self-scratching. While being attended to on the dermatological ward, she was also examined by psychiatrist MK-K, who reported that her "orientation to time, place and person, and also memory and cognition is intact". The patient was communicative, sometimes even garrulous, when describing the course of her disease. She reported many family problems (she was separated from her alcoholic husband but; was still living in the same flat as him because of financial difficulties). She had proprioceptive hallucinations of a "double back" tingling sensation, caused, so she claimed, by viruses and infected fluid flowing through her whole body to her heart. The patient had no insight as to her real condition. She had retired because of her health problems and spent most of her spare time consulting numerous specialists. Her sleep was disturbed as a result of her sleeping in a sitting position in order "not to have her head flooded by internal pus". MK-K also noticed a high level of anxiety, although there were no mood disturbances and the affect was appropriate. She diagnosed the patient with somatic-type delusional disorder (caused by the conversion of emotional tension into somatic symptoms and exacerbated by the menopause) and administered sulpiride at a dose of 200 mg/day. (Kepska et al., 2011, p.193)

Thought insertion- The patient believes that someone else deliberately puts thoughts into the patient’s mind. Sometimes the patient also reports that these thoughts can

force them to do unwanted actions. Also, sometimes patients claim that they know the identity of the person who is doing such insertions. Case report from (Area, 2003):

> The patient was a 25-year-old man with a 1-year history of poor concentration and depressive mood. Apart from an unconfirmed history of fetal distress and epileptic seizures in his childhood, his past medical and psychiatric history, as well as his family history, were unremarkable.
>
> A year before our first interview he started complaining of sadness, poor concentration and "tension in his head", "as if it were to explode". He started to lose confidence, feeling that he was unable to complete his work. By that time he noticed increasing difficulty in thinking, and was suffering from occasional thought blocking and thought withdrawal.
>
> After a few months he started to experience his thoughts "as if they belong to another person". Interestingly, he tried to consciously avoid this sensation by speaking aloud and formulating his inner speech as a dialogue of questions and answers, of which the patient gave us this example (the patient is speaking aloud to himself):
>
> -"Where should I cross the road?
>
> -I should cross by the zebra crossing."
>
> Three months later, however, these "dialogues" evolved into an inner speech with a complicated structure. The patient gave us this example:
>
> "For example, I think without speaking:
>
> -Shouldn't I know where I ought to cross the road?
>
> And the answer comes to my mind quickly:
>
> -I should know that I ought to cross by the zebra crossing…
>
> But now there's no need for speech…The answer comes to me automatically…"
>
> These phenomena were distressing for the patient who felt very embarrassed by those "dialogues".
>
> Initially he was concerned about the perspective of looking for a job outside of the family farm, fearing that he was going to be rejected if people noticed that he had to speak aloud what he was actually thinking. Furthermore, 3 months later, when the phenomenon became an automatic inner dialogue, the patient suffered an important retardation of thinking that impaired his working

performance and made it impossible for him to look for a job on the job market. (Area et al., 2003, pp.129-130)

With the purpose of explaining the diversity of delusions, different theories have been developed. Empirical tests have been devised in order to see if there is a specific reasoning ability that goes wrong in delusional patients. So far, tests have found that there is no specific difference in reasoning skill that could explain delusions. In fact some people with delusions can be defined as being very intelligent, and patients do not seem to perform worse than the average population in reasoning skills. The only test that has proven that delusional people have different skills than the average population is Huq, Garety and Helmsley's (1988) probabilistic study in which patients performed *better* than the average population, showing a tendency to jump to conclusions. The reason why jumping to conclusions is considered "better" in this specific study is because, in this type of Bayesian task, the deluded subjects predict the correct result with less tries (less evidence) than the normal control group. Therefore, this study shows that the patients do better in probabilistic tests because they take less time in obtaining the desired result.

III. Scope of the thesis

The aim of this research is to examine the philosophical debate about the nature of delusions. In order to do that I will draw a conceptual map of the current philosophical accounts: I will show how these views relate to each other by looking at their philosophical assumptions. My analysis will examine examples of and arguments advanced for the main philosophical approaches to delusions.

My intention in this research is not to fit delusions into one of these categories but to shed some light on the underlying assumptions which drive the different competing views which often seem to talk past one another. Therefore, by drawing a conceptual map of these philosophical assumptions, I will be able to shed light on the debate in order to make a diagnosis about the nature of delusions.

This type of research may have direct clinical consequences for the management of delusions. It may also shed light on the nature of beliefs and other propositional attitudes and on processes of belief formation in the normal population.

The aim and importance of discussing delusions is to provide a framework in which to probe our most entrenched concepts, such as the concept of "belief", "rationality", "experience", "abnormal", and many others that we take as given and that form part of the everyday for clinicians, psychiatrists and psychologists.

My suggestion is that there is a general lesson to be learned from the different accounts of delusions: that is, that they are very baffling. Despite the brief characterisations of Capgras and Cotard, they remain very difficult states to understand. What would it be to think that one were dead? What would it be to deny all the apparent facts about one's partner? It seems impossible to understand what is it like to have one of these delusions. That is the lesson from the examples: delusions present a challenge for understanding. At the same time, if they are mental states, as they also appear, understandability is what one would expect.

Second, this thesis attempts to chart attempts made by philosophers to tackle this puzzling feature of delusions through philosophical theories of the nature of delusions. The hope seems to be that by saying what sort of state a delusion is, light can be shed on their status as understandable or not.

Thirdly, the thesis argues that none of the various theories about the nature of delusions succeeds. All accounts face general problems. But my more specific aim is an attempt to show why this is so: I argue that there is something problematic about the form of analysis offered so far. In the final chapter, the latter form will be compared to a specific form of argument called "the likeness argument", criticised by Pickering (2003). The general issue is that all the accounts attempt to shed light on the problem by comparing delusions to more familiar kinds of mental states, however, this faces the difficulty of making delusions too easily understandable. Retreating from this by defining more bespoke, less familiar mental states threatens to make the accounts ad hoc and unilluminating.

IV. Criteria for the accounts' selection

Given that there are many distinct accounts of delusion, and not all of them could figure in this work, I would like to explain my reasons for selecting the ones that appear throughout this thesis. The views that I discuss in the different chapters fulfil two

criteria: 1) they illustrate the categories of my taxonomy, and 2) they are considered either highly influential or very interesting.

Regarding 1), I offer a general taxonomy by classifying accounts of delusions as doxastic, antidoxastic, phenomenological, or meaning-related. I judge it appropriate to discuss delusions in this way because the various accounts fall into general areas; it seems intuitive that delusions are beliefs, hence doxasticism was picked as one of the categories; however, it is also the case that delusions do not seem entirely like beliefs, so, antidoxasticism is its necessary complement. Also, there is a whole other tradition, phenomenology, which does not take delusions as merely cognitive states. Another interesting position is, not just what kind of state delusions are, but what patients actually mean by their utterance. In other words, I chose these four different areas because they enable a fundamental and encompassing way of analysing delusions, and particular accounts should be illustrative of these, more general, frameworks.

About 2), I picked up the authors of specific accounts that were either the most influential or the most interesting. For example, Louis Sass is massively influential amongst philosophers; and John Campbell is considered, arguably, one of the brightest philosophers of his generation. Andy Egan's account is an unusual and very clever approach to the problem of delusions, whilst Tim Bayne and Elisabeth Pacherie are influential in the debate by proposing their own account and discussing other views in interesting ways.

Although I had to leave many other appealing accounts and general frameworks out of this work, in general, I consider that the ones that appear here are very illustrative for the purpose of showing the status of the debate.

V. Summary of the chapters

The first chapter of this thesis is an overview of the central concepts developed by Karl Jaspers regarding the idea of delusion, which are common to contemporary accounts of the phenomenon. Through this chapter I explain the concepts necessary to reach two different definitions of delusion: one via its features, and the other one by being ununderstandable.

The second and third chapters are a critical analysis of two opposing philosophical views: antidoxastic and doxastic accounts of delusion. Antidoxastic views propose that delusions are better understood as a propositional state different from beliefs. In those chapters, I argue that the advantages of both antidoxastic and doxastic views are limited due to their ad hoc nature. The alternative, the doxastic view, argues that delusions are beliefs despite their abnormalities and having little common ground with our everyday beliefs. In this chapter I analyse the concept of belief and argue that it is flawed in itself, undermining doxastic views.

The fourth chapter considers phenomenological accounts as an alternative to the debate presented in the two previous chapters. Phenomenology presents itself as an appealing account due to its encompassing nature which takes into account the experience of the person itself. In this chapter I argue, however, that the understanding one gets through phenomenology is limited. This, I argue, is due to the flexible nature of the concepts involved in a phenomenological account; as a consequence, many of the important puzzles in delusion remain untouched.

The fifth chapter concerns, through the accounts given by Berrios, Sass, and Campbell, the meaning of utterances in delusion. Although the propositions of a deluded patient seem to be like the ones of a non-deluded subject, their meaning is not straightforward. Accounts that try to interpret this anomaly have not provided a satisfactory answer.

The final chapter is a diagnosis of why the previously discussed accounts have failed to capture the nature of delusion. My main argument is that the form of the argument used, labelled “the likeness argument” by Pickering, is fundamentally flawed, therefore affecting all possible agreement in this debate.

CHAPTER I - KARL JASPERS AND THE CONCEPT OF DELUSION

I. Introduction

The aim of this chapter is to set the scene for the rest of the thesis by examining Jaspers' discussion of delusion. This is so, firstly, because Jaspers is discussed on several occasions by different authors analysed in this thesis. Secondly, Jaspers' concepts are the foundation of many contemporary ideas and the root of several assumptions that permeate the contemporary debate. But more importantly, Jaspers' definition of delusions sets the bar against which many features are tested.

In his book *General Psychopathology*, Jaspers (1913), offers two definitions of delusions. One has continued to be influential through its modification in definitions of delusion in versions of the *Diagnostic and Statistical Manual* (American Psychiatric Association, 2013). The other, more relevant to later chapters here, defines delusions via the idea that they are "ununderstandable". Given that Jaspers also suggests that understanding and understandability lie at the heart of the mental and are central to psychiatry this suggests an apparent contradiction. The most important symptom in psychiatry is not accessible using Jaspers' preferred method. I will suggest that attempting to show how delusions can be understood, despite Jaspers' claims, has been the main agenda of recent philosophical work on the ontology of delusion.

In the first section of this chapter I analyse a central concept for Jaspers' definition of delusions: understanding. Understanding, according to Jaspers, contrasts with explanation and can take two different forms: genetic and static. The former, genetic understanding, is what constitutes the concept of empathy, which will be analysed in its own section here. Jaspers thought that mental phenomena could be grasped by a certain way of looking at the patient's utterances, and he calls this method "empathy". Empathy is important here because Jaspers thought that delusion is beyond the reach of empathic understanding.

The next concept I will discuss is "delusional atmosphere". This is related to what psychiatrists today would consider as a premorbid state and, therefore, is very close to the concept of being fully delusional, which Jaspers would have labelled as being in a "delusion proper". After that, I discuss some related concepts, such as delusion-like ideas, along with primary and secondary delusions.

Finally, I look in detail at both branches of Jaspers' definition of delusions. The first definition informs what we know today as the DSM definition of delusion. It enumerates the characteristics that clinicians look for in delusional patients. The second stance deems delusions a phenomenon that, unlike all other mental afflictions, has the character of being ununderstandable.

II. Understanding in Jaspers

a) Background

In order to grasp the concept of understanding in Jaspers' work, it is useful to consider the philosophical background with which he was working. Karl Jaspers (1883-1969) was one of the most influential philosophers in psychopathology of the twentieth century. He was highly influenced by the works of Husserl and Heidegger, and thus his own work has a strong phenomenological element.

In Jaspers' time, what psychiatrists aimed for was to find the specific area in the brain that caused syndromes or mental illnesses. In this framework, Kraepelin was one of the most prominent influences; he explained mental disease in terms of its organic symptoms. Other individuals, such as Griesinger, thought that there was nothing more to mental disease than cerebral malfunction "Mental diseases are brain diseases"(Griesinger, 1845). It is in opposition to this conceptual framework that Jaspers proposed a more encompassing philosophy that took into account the individual and his psychic life. Jaspers recognizes possible environmental influences and, remarkably, the experience itself suffered by the patient, which was uncommon in his time.

Against Griesinger's assertion that "mental diseases are brain diseases" Jaspers writes:

> This declaration [from Griesinger] is as dogmatic as its negation would be. Let us clarify the situation once more. In some cases we find connections between physical and psychic changes taking place in such a way that the psychic events can be regarded with certainty as consequences. Further, we know that in general no psychic event exists without the precondition of some physical basis. There are no 'ghosts.' *But we do not know a single physical event in the brain which could be considered the identical counterpart of any morbid psychic event.*

We only know conditioning factors for the psychic life; we never know the cause of the psychic event, only a cause. (Jaspers, 1913, p.459 my italics)

Therefore, in this context, Jaspers aims to outline a concept of mental disease clearly different from the previous existing ones. On the other hand, Jaspers gives a positive account of what psychopathology should be concerned about. The concepts of understanding, and more specifically empathy, which I will explain further on in this chapter, fulfil this necessity.

b) Jaspers' definition of understanding in *General Psychopathology*

Jaspers has a particular way of defining the concept of "understanding", drawing on a phenomenological tradition. He distinguishes two sorts of understanding: Understanding the mental phenomenon as an experience (as most phenomenologists of his time did), and understanding through empathy. This dual view of understanding is unique to Jaspers. Understanding, in both variants (static and genetic), has the following characteristics:

- The parts of the psychic life where understanding is not possible fall under the term "interpretation"[2], that is, just to make sense of the particular tangible facts. Therefore, it might be possible to interpret a bodily disease, such as heart failure, but this is not what it is done in the case of mental disease.
- Although it can be useful to make some sense of the sets of characteristics of the disease, interpretation *does not* lead to understanding.
- Understanding requires the intuition and abilities of the clinician so that the psychic features of the patient can emerge. To Jaspers, understanding is the mark of the mental.
- The limit of understanding is the ununderstandable (Cfr. Jaspers, 1913, p.27-28).

[2] *Interpretation,* to Jaspers, is what can be achieved through *tangible facts* (verbal contents, cultural factors, people's acts, ways of life, and expressive gestures). "All such objective data, however, are always incomplete and our understanding of *any particular , real event* has to remain more or less an *interpretation* which only in a few cases reaches any relatively high degree of complete and convincing objectivity" (Jaspers, 1913, p.303). Understanding, on the other hand, is perceived through subjective facts, it is something that emerges from the phenomenon.

Summarizing: Jaspers' understanding is the way to reach other people's inner states. It is described as follows by Fulford et al:

> "Understanding" (*Verstehen*) for Jaspers is the route to other people's inner mental states. It underpins the ability to "read" their motives and subjective meanings from their actions and speech. (Fulford et al, 2006, p. 169)

It is achieved through personal intuition and it reveals meaningful psychic connections. However, understanding is not a causal explanation.

According to Jaspers, the fundamental difference between understanding and explanation thus has two dimensions: the epistemic and the ontological. In the epistemic dimension, "explaining" requires a *repeated experience* which allows us to make relevant theories. That is how one comes to know about the causal connections of explanation. In understanding, an analysis of a *particular case* is needed in which *personal intuition* is required for every case. That is, every case is different, and an approach that worked for a specific case might not work for another. Personal criterion plays a role in understanding.

In the *ontological* distinction between understanding and explanation, the difference consists in the following: Understanding provides us with knowledge of *meaningful psychic connections*, whilst explaining provides, instead, *rules of causality* (Cfr. Jaspers, 1913, p.304). That is the difference in what is known. I will return to the significance of this shortly.

c) Genetic understanding and static understanding

Within the more general concept of understanding Jaspers proposes a distinction between static and genetic understanding. This is relevant to the concept of delusion because genetic understanding is equivalent to the concept of *empathy*, which is the tool that Jaspers uses in order to analyse any mental state in order to rationalize other people. As I will describe at the end of this chapter, the failure of empathy is the defining feature of primary delusions.

Static understanding is defined by Jaspers as follows:

> The *static mode* denotes the presentation to oneself of psychic states, the objectifying to oneself of psychic qualities, and we shall exercise this kind of understanding when we come to the chapters on **phenomenology** and the psychology of expression, etc. (Jaspers, 1913, p.17 bold emphasis added)

Therefore, static understanding consists mainly in applying phenomenology in general to generate an insight into the very possibility of types of psychopathological symptoms. One might think of this as *what is it like* to have a certain sort of experience.
Genetic understanding, in contrast, is defined by Jaspers as follows:

> [W]e shall occupy ourselves with the *genetic mode*, that of **empathy**, of perceiving of psychic connections and the emergence of one psychic phenomenon from another. (Jaspers, 1913, p.17 bold emphasis added)

In other words, static understanding describes the phenomenon using phenomenology, but its reach is limited to the enumeration of the symptoms. In contrast, genetic understanding will provide an understanding of psychic, meaningful connections. Thornton (2007) makes a summary of this relationship:

> Thus, the relationship between static and genetic understanding is like this. The former articulates and vividly presents what it is like, for example, to have a sudden realisation or what it is like to be in a state of happiness. It makes these kinds of state clear for further inquiry prior to the imposition of psychological theory. Genetic understanding adds to this the connection of how one state arises –ideally and typically- from the other. Such connections are shared empathically by psychological subjects including psychiatrists and their patients. (Thornton, 2007, p.96)

Genetic understanding is renamed by Jaspers as empathy in later works. Because Jaspers argues that delusions are characterised by empathic inaccessibility, I will ignore static understanding in what follows and turn to characterise genetic understanding, or empathy, in the next section.

III. Empathy

Empathy, that is, genetic understanding, is characterized by Jaspers as follows:

> "Genetic understanding" [is] the understanding of the meaningful connections between one psychic experience and another, the "emergence of the psychic from the psychic" (Jaspers, 1968, p.1322).
>
> [Subjective symptoms] cannot be perceived by the sense organs, but have to be grasped by transferring oneself, so to say, into other individual's psyche; that is, by *empathy*. They can only become an inner reality for the observer by his participating in the other person's experiences, not by an intellectual effort. Subjective symptoms include all these emotions and inner processes, such as fear, sorrow, joy, which we feel we can grasp immediately from their physical concomitants. (Jaspers, 1968, p.1313)

What does Jaspers mean by empathy being constituted by "participating in the other person's experiences" (as I quoted above)? Evidently, this description should not be taken literally: to have empathy for someone who feels depressed does not necessarily mean to feel depressed oneself. So, what did Jaspers truly mean? He explains this further on in his essay:

> The first important differentiation was made by Simnel, who showed the difference between the understanding of what has been said from understanding the speaker. When the contents of thoughts emerge one from another in accordance with the rules of logic, we understand the connections rationally. But if we understand the content of the thoughts as they have arisen out of the moods, wishes, and fears of the person who thought them, we understand the connections psychologically or empathetically. Only the latter can be called 'psychological understanding'. Rational understanding always only enables us to say that a certain rational complex, something which can be understood without any psychology whatever, was the content of a mind; empathic understanding, on the other hand, leads us into the psychic

> connections themselves. Whereas the rational understanding is only an aid to psychology, empathic understanding is psychology itself. (Jaspers, 1913a, p.83)

However, what he means by empathy is vague. It is a notion that permeates all of Jaspers' *General Psychopathology*, but that can be understood in different ways by different phenomenologists. Empathy is more like an exercise where the possibility of clarification comes from the dialogue between the patient and their interaction.

Jaspers' concept of empathy presents two fundamental problems due to its ambiguous nature:

a) It is too general and too vague to put it into a context where it can be falsified or confirmed by empirical experience. For example, one cannot put to the test whether a clinician used empathy or just a mere rationalization in order to diagnose a patient. If empathy just means something like "examining the feelings of the person", it stops being a distinct method to analyse mental states, as Jaspers claimed.
b) Or it becomes so specific that it loses some of its original properties, therefore, it doesn't reach its original goals, and becomes unimportant. That is, if empathy becomes just tracing specific features, such as "how many days have you felt depressed?", as is the case in some semi-structured interviews, that would risk ruling out some forms of empathic understanding, therefore, its importance is reduced to mere *explanation*.

In other words, empathy is difficult to test empirically either way because of its ambiguous nature. However, Jaspers did not believe that empathy was something opposed to other empirical descriptions of psychopathological issues. To him, both could be complementary in the sense that his view could contribute to a better understanding of purely empirical views, and vice versa.

IV. Hoerl's distinction of explanation and understanding

I have discussed in the previous sections how understanding, for Jaspers, is not the same as an explanation. Understanding, also, is not causal. Jaspers characterizes understanding with these two features because he does not want to fall into a purely reductionist account in which mental phenomena are explained solely as brain reactions

which are properly subject to causal explanation. However, it is difficult to understand what Jaspers meant by understanding given that it is not causal, nor an explanation. However, Hoerl has a plausible response to what could be the case given Jaspers' restrictions to the concept.

In his (2013) paper, Hoerl wants to defend Jaspers' intuition that explanation and understanding are conceptually different. However, this is more complicated than it initially seems because there is, in principle, a dilemma that Jaspers' account of genetic understanding faces: if genetic understanding is compatible with tracking causes then its results are indistinguishable from causal explanation; if it is not, then, what kind of information can understanding give about a person's mental life?

Adopting the first part of the dilemma, seen from a Humean perspective, both explanation and understanding are similar because they appeal to the same notion of causation in which regularities seem to give birth to laws.

Still, there is another problem from Jaspers' definition of understanding. Jaspers claims that understanding leads to "meaningful psychic connections" but that these connections are not causal. However, if this were the case, it is unclear what knowledge we could actually gain from them. Another contradiction that Hoerl highlights is that Jaspers says that "psychic events *emerge* out of each other in a way which we can understand" (Jaspers, 1913, p.302 italics added). The drawback here is that emergence of this sort necessarily implies some sort of causation, although Jaspers himself tries to deny that.

Hoerl, in response to these problems, suggests a middle ground position that relies on the idea that understanding traces some sort of causality but nevertheless differs from explanation. He relates his position to two of the most influential competing theories of the role of understanding: Campbell's emphasis on causation alone versus Davidson's and Dennett's theories of interpretation based on rationality. On the one hand we have Dennett's and Davidson's claim that understanding by ascribing propositional attitudes can be achieved via making *rational sense* of beliefs, desires, etc. (though, as we will see in later chapters, in the case of delusions, this sort of rational stance falls apart because delusions are either rational responses to a bizarre experience or people with delusions do not actually believe in their delusion). This leaves us with a concept of understanding which is rational and intelligible whether or not it is also causal. Campbell's (2008) discussion of causation suggests that we should, instead, think of causal relationships

without the concept of understanding, for instance, as what can be said to cause a mental state. He denies the concept of "levels of explanation" in psychiatry. Levels of explanation can mean two things:

> There are two dominant approaches to the idea of levels of explanation: ontological and epistemological. (…)
> The ontological view is part of a traditional reductionist picture of the world. On this picture, sciences of the mind, such as psychiatry and psychology, can in principle be reduced to biology (which might be construed as physiology or evolutionary biology), biology to chemistry, and chemistry to physics (…)
> There is, however, another and quite different approach to levels of explanation, which has been influential. (…)
> This hierarchy does not concern different ontological levels but rather different ways of understanding the same ontology. The highest, most abstract level concerns the function of a system. It might be carried out by a variety of different algorithms at the middle level. Finally, the same algorithm might be realized in different physical ways at the lowest level. Thus, the higher levels are multiply realizable by lower levels. Determining the computational level is a matter of determining the goals of a system independently of its physical or neurological properties. (Thornton, 2015, p.2)

Given such types of levels of explanation, Campbell argues that neither can work for psychiatry. Campbell thinks that there is no reason to hold that the world makes sense at particular levels of explanation.

> We naturally seek a certain kind of intelligibility in nature; we naturally try to find explanations that will show the world to conform to reason, to behave as it ought. Hume's point is that there are no such intelligible connections to be found (…) We are prone to relapse, to think that after all we must be able to find intelligibility in the world. This tendency survives, I suspect, in the idea of 'levels of explanation'. The idea is that within certain levels of explanation, we will find a particular kind of intelligibility. [T]he lesson from Hume is that there is no more

> to causation than arbitrary connections between independent variables of cause and effect. (Campbell, 2008, p.201)

There is brute causation (for which he does adopt a Humean law-like approach) and that is all and it can operate in any way. So there is no level of explanation or kind of explanation that we can call 'understanding' and contrast with the rest of explanation. But the problem with that is that it does not tell us what aside from causal connections understanding might give us.

Recall first the dilemma for Jaspers: how can understanding be different from explanation but still tell us about how mental states arise from one another, which Jaspers thinks it can? Jaspers suggests that understanding is not causal; but then, it is difficult to know how this type of understanding can provide us with insight about the world. Hoerl suggests that we can contrast particular forms of causal connection and lawlike forms. So understanding can give us *one off* causal connections. The question is whether that is enough for understanding. Is that all there is to it? If so it would fit Campbell's minimal view. According to Hoerl (2013), both Dennett and Campbell generate an apparent paradox for Jaspers. Hoerl's proposal is that understanding is different from mere rationalizing, rather, there are some a priori constraints on what may count as rational thought.

> On one possible reading, what Jaspers might be seen to be drawing attention to in these passages are features of human psychology that are simply left out by the idea of the mind as governed purely by principles of rationality, but that, in turn, have some bearing on the extent to which the subject can be rational (…)So we might conjecture that understanding, for Jaspers, in so far as it is to be contrasted with or go beyond mere rationalizing, is in fact concerned with features of psychology of the latter type, that are in some sense preconditions for, and constraints on, rational thought. In this category we might further include, for instance, the conditions under which certain thoughts occur to us in the first place, or come to occupy us, or indeed become difficult to shake off despite some evidence to the contrary. (Hoerl, 2013, p118)

This still leaves the earlier dilemma (between understanding collapsing into causal explanation or failing to provide insight into the meaningful connections between mental states). A plausible response from Hoerl is to embrace the first part of the dilemma but argue that Jaspers could ascribe a singular epistemic causation to mental events that is not necessarily a Humean lawlike relation. He describes the latter as follows: "We might thus say that Jaspers subscribes to a form of *epistemic particularism* regarding understanding" (Hoerl, 2013, p.108). With regards to its epistemic implications, to Jaspers, every particular case is something *self-evident* that cannot be broken down any further into more basic or fundamental pieces. According to Hoerl (2013), Jaspers' statement can be better apprehended by Anscombe's concept of singularism about causation, given the fact that the former wants to distinguish causation in the form of inference from the process of "understanding":

> […] only understanding makes manifest what causation between one element of a person's mental life and another ultimately consists in. In so far as there are true generalizations or laws in psychology at all, these obtain in virtue of the kinds of *singular causal connections* that understanding makes manifest. The obtaining of such generalizations or laws is not what makes it the case that one psychic event causes another on any particular occasion. (Hoerl, 2013, p.111)

In this way one can have a concept of genetic understanding which is not a form of Humean causation but that might still be somehow intelligible. This is useful because genetic understanding is the relevant form of understanding in Jaspers for considering the concept of delusion. Hoerl's account of singular causal connections retains the characteristics ascribed to Jaspers' original description whilst bringing a modern view to the concept of understanding.

V. Concepts previous to the definition of delusion: delusional atmosphere, delusion-like ideas, and the distinction between primary and secondary delusions

a) Delusional atmosphere

Before discussing how Jaspers deploys ununderstanding as a mark of delusions, I will first set out some notions related to, but distinct from, primary delusions according to him. These are the concepts of delusional atmosphere, and delusion-like ideas and then the distinction between primary and secondary delusions.

Delusional atmosphere is a phenomenon in which the patient suspects that there is something different in her own reality. It is the point at which the patient starts thinking that something in the world or, as it were, the atmosphere, is wrong. At this stage the patient has not yet formed the bizarre belief. Delusional atmosphere is the earliest signal of the imminence of delusion. It is described by Jaspers as follows:

> Patients feel uncanny and that there is something suspicious afoot. Everything gets a new meaning. The environment is somehow different –not to a gross degree- perception is unaltered in itself but there is some change which envelops everything with a subtle, pervasive and strangely uncertain light. A living-room which formerly was felt as neutral or friendly now becomes dominated by some indefinable atmosphere. Something seems in the air which the patient cannot account for, a distrustful, uncomfortable, uncanny tension invades him. (Jaspers, 1913, p.98)

German E. Berrios (1991) observes that Jaspers based his notion of delusional atmosphere in previous similar concepts from the nineteenth century. These notions come from (Hagen, 1870), (Störring, 1939), (Conrad, 1958), and (Llopis, 1969), amongst others.

According to Jaspers, delusional judgement is a different, secondary stage of the whole experience, in the sense that delusions are *first* presented via "delusional atmosphere", or "delusional mood". However, this has to be read carefully, since Jaspers is not saying that delusional atmosphere causes delusions to arise. The latter is not an aetiological

claim: “Rather, primary delusions become intelligible possibilities in the context of a shift in the overall structure of the experience” (Ratcliffe, 2013, p.231). The sense in which this could be interpreted is the following: delusional atmosphere precedes (in time) delusion proper; that is, delusional atmosphere is exactly what happens before the delusion, as a fixed thought, is present. However, this is not a causal claim; it is not that the delusional atmosphere causes the delusion proper to arise, rather, both are parts of the same phenomenon, in the same sense that flu is preceded (but not caused) by a sore throat, delusional atmosphere precedes (but does not cause) the delusional judgement. Although the use of the word “atmosphere” suggests something vague, as Jaspers admits, it has to do with a meaning that can be attached to the symptoms of every delusional case. It reflects the suffering of the patient, and can only be relieved when the patient reaches a definite delusional idea, because “no dread is worse than the danger of the unknown” (Jaspers, 1913, p.98).

Jaspers argues that the sense of reality that has changed in delusional atmosphere is not exclusive to attitudes of belief, since this change is something more integral to the perceptual experience. In other words, as Ratcliffe puts it, what is “there” depends on what one believes to “be real” a priori. From a phenomenological point of view, this does not have to do with being a belief or not. So, the experience would affect not only belief states, but other states as well: “So other kinds of intentional states, such as imagining, remembering, anticipating, doubting, and so forth (…) equally depend upon the experienced sense of reality” (Ratcliffe, 2013, p.232). This matter becomes relevant in later chapters, because it undercuts the motivation for the debate between doxastic and antidoxastic theories of delusion. The question of whether the delusion is a doxastic or an antidoxastic state is a non-question to Jaspers and many more recent phenomenologists.

Jaspers suggests that delusional atmosphere is like living in an “immediate and shifting present” (Jaspers, 1913, p.104). This agrees with many of the patients’ descriptions of their experiences. One illustrative example can be drawn from a patient called Sophie, described in Sass’ (2014) paper:

> If I no longer believe in gravity, it’s not that I fail to anticipate something when I don’t expect an apple to fall from the tree, but that I simply think that the apple

> could just as easily float or fly and therefore have no reason to anticipate it falling. (Sass, 2014, p.144)

And also, one can witness the experience of a shifting metaphysical framework:

> I cannot count the number of times I've been told "but Sophie, X is impossible" and all I ever want to say in response is "yes, I am perfectly capable of appreciating why you think X is impossible, but your conceptual or metaphysical constraints are simply not mine". (Sass, 2014, p.144)

As Ratcliffe (2013, p.233) puts it, this new set of ideas is not anchored any more to the public realm where it can be revised, discussed, debated or questioned.

Jaspers emphasises another feature of delusion: that the *content* of the delusional atmosphere is only accidental. This means that, although the subject might say that he is dead, that the delusion takes this specific form is not what truly matters in delusion. To Jaspers, the patient might as well have said that he is Napoleon, or that he is being persecuted, and this would make no substantial difference to the overall experience of the delusion. This differs greatly from Freudian, psychoanalytic views in which the content of the delusion points to its aetiology. Jaspers is so radical in his assertion about the content that he goes so far as to suggest that even the patient might not have a total grasp on the content of the delusion: "[H]owever, it is certainly possible to wonder whether the patients have found any content adequate for their actual experience" (Jaspers, 1913, p.99). To Jaspers, the correct way in which one could look at this experience would be through the analysis of the patient's feelings and sensations, and not whether he thinks he is Napoleon.

But what is the "experience of reality" according to Jaspers? It is a phenomenon that cannot be defined, deduced or compared. It is a primary phenomenon itself and its characteristics are primary in themselves as well. We can only describe it *indirectly*. To Jaspers, the concept of "reality" entails the following (Cfr, Jaspers, 1913, p.94):

1) What is real is what we concretely perceive - By this, Jaspers does not mean that what is real is the information given by any sensory organ in particular, but *the sense* of what is actually collected by this organ.

2) It lies in the awareness of Being. The real, existing entity is the thinking-being that is aware of its own thinking in a Cartesian way. It exists even within phenomena as "derealisation".
3) What is real is what resists us. Reality has at its core the *interpretation* of meanings, events and situations.

As we can see from this kind of definition, reality does not necessarily adhere to a description of a shared world, full of objects. By these guidelines, reality is constituted by a subjective, personal, interpreted world which is not necessarily shared by others. We can see in Jaspers' definition the need for other minds to exist and be understandable according to his three guidelines above. This is the root of many contemporary phenomenological accounts, such as "multiple realities" stances.

Delusional atmosphere is a state which contemporary clinicians would consider as already being delusional, or perhaps, a pre-morbid state. To Jaspers these definitions would make no sense, because what one is actually describing is *the patient's experience*, and this can come in different stages, but it would still be the same broader experience.

To Jaspers, the key factor is the transformation of the patient's experience, but as he himself acknowledges, this is precisely the part that is hard to grasp. The concepts work together in the following way: The delusional atmosphere is produced by this sense of having a distorted experience of reality. The delusional atmosphere precedes the actual delusion, which we will call delusion proper.

b) The distinction between delusion proper and delusion-like ideas

As a second piece of preliminary business, before I discuss delusions proper below, there is another phenomenon Jaspers discusses: delusion-like ideas.

Delusion-like ideas are not delusions; they only look like delusions because they are implausible or false. These are ideas which emerge from comprehensible sources. Examples of such are depression, mania, and overvalued ideas (Cfr. Jaspers, 1913, p. 107):

> One group [the group of delusion like ideas] *emerges understandably* from preceding affects, from shattering, mortifying, guilt-provoking or other such

> experiences, from false-perception or from the experience of derealisation in states of altered consciousness, etc. (Jaspers, 1913, p.96)

> The term delusion-like ideas is reserved by us for those so-called "delusions" that emerge comprehensibly from other psychic events and which can be tracked back psychologically to certain affects, drives, desires and fears. We have no need here to invoke some personality change but on the contrary can fully understand the phenomenon on the basis of the permanent constitution of personality or some transient emotional state.
> Such overvalued ideas [that is, some instantiations of delusion-like ideas] must be clearly differentiated from delusions proper. They are isolated notions that develop comprehensibly out of a given personality and situation. (Jaspers, 1913, p.107)

Jaspers therefore denies that such comprehensible states should be called delusions:

> The term delusion should properly only be given to those delusions which go back to primary pathological experiences as their source, and which demand for their explanation a change in the personality. (Jaspers, 1913, p.106)

Jaspers explains that delusion-like ideas are different from delusions proper in the following way. The former are developed in similar ways to our everyday, normal experiences, and are understandable. Examples of such delusion-like ideas are over-valued ideas, depression, and transient states that are due to false perception (Cfr. Jaspers, 1913, p.107). Delusions proper are different because their epistemic formation is unknown, and are therefore, ununderstandable[3]. They are a "transformation in our total awareness of reality" (Jaspers, 1913, p.95). Jaspers maintains that delusions proper are an experience which is beyond understanding. They cannot be tracked back to personality or to a one-time transient alteration:

[3] That delusions proper are ununderstandable is analysed in detail in the next section

> If we try to get some closer understanding of these primary experiences of delusion, we soon find we cannot really appreciate these quite alien modes of experience. They remain largely incomprehensible, unreal and beyond understanding. (Jaspers, 1913, p.98)

Delusions proper are different from hallucination and delusion-like ideas because the former contain an idea, a judgment, which still arises even against strong contradicting evidence. This is not the case in delusion-like ideas. Proper delusions can only be understood in terms of a disease process:

> Delusions proper are the vague crystallisations of blurred delusional experiences and diffuse, perplexing self-references which cannot be sufficiently understood in terms of the personality or the situation; they are much more the symptoms of a disease process that can be identified by the presence of other symptoms as well. (Jaspers, 1913, p.107)

In sum, delusion-like ideas and delusion proper are two distinct symptoms of two different pathologies. Delusion-like ideas are comprehensible, possibly transient symptoms; on the other hand, delusions proper are the root of an ununderstandable disease. Only the latter should be called delusions.
In the literature, some philosophers blur two different distinctions: that between delusions proper and delusion-like ideas, and that between primary and secondary delusions. I will dedicate the next section to discuss that view.

c) Primary and secondary delusions

Jaspers makes a distinction between primary and secondary delusions. The former are ununderstandable, where delusional perception does not "involve misinterpretation of perceptions and experiences" (Kraus, 2014, p.110). In *primary* delusions what is important is the disturbance of the ego and the phenomenological experience. In this primary state, the patient tries to understand his strange experience given his delusional mood, and has a vivid experience of the delusional content.

The latter, secondary state of delusion, is similar to the DSM definition of delusions as "erroneous beliefs" or false judgements. Each of these two definitions gives birth to different contemporary traditions in philosophy of psychopathology. The notion of delusion as the ununderstandable is the basis of phenomenological contemporary traditions. By contrast, most analytic, functionalist contemporary traditions took Jaspers' other definition of delusions, and its influence can be seen in modern DSMs.

There are some claims in the literature (e.g. Thornton, 2007) that primary delusion is equivalent to delusions proper, whilst secondary delusion is equivalent to delusion-like ideas. These assertions might be motivated by Jaspers' explanation that delusion-like ideas are comprehensible whilst primary delusions are not understandable. However, Jaspers' actual distinction goes like this: delusions proper are ununderstandable. But this is only the case in its primary stage of delusions proper. Regarding the other case, delusion-like ideas, I have explained in the previous section that these are not delusions at all. The fact that they are comprehensible does not mean they are the counterpart of the ununderstandable. The distinction between primary and secondary delusions is only to differentiate the stages of what Jaspers calls delusions proper. In other words, delusion-like ideas are not delusions, whilst delusion proper has two stages, that is, primary and secondary.

VI. Two definitions of delusions in Jaspers

In this section I present the two definitions of delusions in Jaspers' work. The first one describes delusions through their pathological symptoms and corresponds to the secondary stage of delusion, where the judgement has been formed and the features that characterize delusion are more salient. This account was central in the definition of delusion in the current DSM. However, to Jaspers, it was a distraction from the true nature of delusion. The real characterization of delusion, to Jaspers, comes from its primary state, which has the mark of being ununderstandable. Both definitions have an impact in contemporary philosophical accounts. I will start with the definition Jaspers thought less helpful: via its characteristics.

a) Delusions defined via their characteristics

To Jaspers, the definition of delusion via its characteristics is "vague" and beside the point because it is not capable of reflecting the true nature of delusion. For him, characterizing delusions in this way is just a useful tool to recognize the point when the secondary delusion has come into play. Jaspers explains that delusions can be defined via their apparent, external characteristics that come in an undefined degree as follows (Cfr. Jaspers, 1913, p.95-96):

I) They are held with extraordinary conviction and incomparable subjective certainty.

II) There is an imperviousness to other experiences and to compelling counter-arguments.

III) Their content is impossible.

About I), Jaspers says that the convictions sustained by the patient are not like normal convictions because the patient behaves towards them in a certain "ironic" way to start with, and secondly, the patient seems to sustain them in a different quantitative manner to what is seen in normal convictions, although in "undefined degree" with a subjective certainty (Jaspers, 1913, p.80). The absolute certainty in delusions is only related to the delusional idea via the reality of the experience, while the delusional perception is based in the "common sense reality" (Kraus, 2014, p.117-118).

Point II) is regarded by Jaspers as a defining characteristic of schizophrenia, given the fact that he sometimes calls it "specific schizophrenic incorrigibility" (Jaspers, 1913, p.88). Jaspers thinks this delusional characteristic is strongly related to point I), saying again how the reality of the patient has been transformed. He states that, in delusion proper, the schizophrenic experience is incorrigible because of "an alteration of the personality" (Jaspers, 1913, p.88). Jaspers, quoted by Kraus (2014, p.118), says that "we are unable to describe the former [alteration of the personality], let alone to formulate it conceptually; instead we [can] only assume it". The imperviousness to other experiences also comes in an undefined, abnormal degree; this imperviousness is not a simple mistake or bias: "Error can also be incorrigible, but in delusion there is always something beyond normal incorrigibility" (Jaspers, 1913, p.342).

Regarding characteristic III), Jaspers thinks that the content of the delusion, although an accident and not the core of the matter, has to be specific, due to *not* being the result

of an error or a lack of knowledge. It is worth noting that Jaspers means that the delusion is “impossible” only in the sense that it is not coherent with the patient’s temperament and history, not that the delusion is literally impossible (for example, paranoid delusions are not impossible as such; the point is that they do not follow from the patient’s narrative). Influencing many of the contemporary philosophers of delusion, Jaspers stresses that the patient does not fully behave as the content of his delusion would dictate to a normal person: for him, delusion is similar to a “symbol for something quite different” (Jaspers, 1913, p.105) because patients do not act in a way that is consistent with what they say they believe; a paranoid patient does not act like a normal subject that is being watched and chased.

The above definition gave rise to the DSM definition. The question now is how similar it remains regarding the modern DSM. In DSM-5, delusions are defined as follows:

> Delusions are fixed beliefs that are *not amenable to change in light of conflicting evidence*. Their content may include a variety of themes (…) Delusions are deemed bizarre if they are *clearly implausible and not understandable* to same-culture peers and do not derive from ordinary life experiences. (…) Delusions that express a loss of control over mind or body are generally considered to be bizarre; these include the belief that one's thoughts have been "removed" by some outside force (thought withdrawal), that alien thoughts have been put into one's mind (thought insertion), or that one's body or actions are being acted on or manipulated by some outside force (delusions of control). The distinction between a delusion and a strongly held idea is sometimes difficult to make and depends in part on the degree of *conviction* with which the belief is held despite clear or reasonable contradictory evidence regarding its veracity. (American Psychiatric Association, DSM-5, 2013, p.87, my italics)

The concepts of “conviction”, “not amenable to change in light of conflicting evidence” and “clearly implausible” are still the main core of the definition. However, Jaspers thought that this way of defining delusion was not able to capture its true essence:

> If we want to get behind these mere external characteristics of delusion into the psychological nature of delusion, we must distinguish the original experience from

> the judgment based on it, i.e. the delusional contents as presented data from the fixed judgment which is then merely reproduced, disputed, dissimulated as occasion demands. (Jaspers, 1913, p.96)

In other words, Jaspers thought the true characterization was related to the experience, not the judgement or its characteristics. Although the general contemporary approach in the latter defines delusions as "mistaken beliefs", Jaspers disagrees with that over-simplified statement:

> To say simply that delusion is a mistaken idea which is firmly held by the patient and which cannot be corrected gives only a superficial and incorrect answer to the problem. Definition will not dispose of the matter. (Jaspers, 1913, p.93)

Ratcliffe affirms that "it is not the transition from the experience to the belief that resists characterization, but the experience itself." This quote from Ratcliffe raises the following question: Is the "experience itself" the delusion? This doubt is relevant because it concerns the ontology of delusion. For example, an important debate within analytic philosophy of psychiatry concerns whether delusion is a doxastic or non-doxastic mental state but against a background assumption that it is some sort of propositional attitude. By contrast, the phenomenological tradition following Jaspers' wider descriptions suggests that that issue is misplaced because the nature of delusion is a broader constellation of experiences. The questions that Jaspers would ask are different, such as if the phenomenon is understandable (in a Jasperian way) and can be reached by empathy.

To Jaspers, in order to analyse delusions proper, an insight into the patients' reality would be required, because to be deluded implies a transformation in the *awareness of reality*. Therefore, the bizarre incident plays only an accidental, less important, role in the process. It is the change in the (phenomenological) Being that alters the patient's experience, which involves many aspects of the subject as a whole, and therefore, goes beyond the mere bizarre incident.

According to Kraus (2014), although the definition of the DSM IV and V has been heavily based on Jaspers' own description of delusions, he did not agree with the idea of looking at delusions just as mistaken beliefs. Jaspers' objection to this kind of definition is not that it is wrong or right in the first place. This sort of question is beside the point, missing

what the core matter of delusion is, “there is a failure to engage with what is most fundamental to them” (Ratcliffe, 2013, p.230). Jaspers did not necessarily agree with the definition of delusion as a belief or not. He constantly uses both the word “judgement” and “idea” to refer to delusion, whether it is primary or secondary. The concept Jaspers is actually keen to use is the concept of delusion as an *experience*. To Jaspers, this is the central approach to delusion in its many stages and through all of its forms. Summarizing: Jaspers did not think of delusions as mistaken beliefs because he thought of them as a more complex sort of experience.

b) Delusions as the ununderstandable

As I have said, Jaspers set out two contrasting definitions. I have described delusions seen through their clinical characteristics. I now turn to delusions as the ununderstandable, which Jaspers thought was the correct way of seeing the phenomenon due to its rich description as an experience rather than a collection of symptoms. The description of delusion as the ununderstandable is deeply phenomenological, and it makes use of the terms understanding, empathy and experience as I will describe.

Although understanding is the “mark of the mental”, Jaspers suggests that there are some limits to this understanding. These limits are reached when treating patients with delusions proper, as is the case in schizophrenia.

> In psychopathology our genetic understanding (or perception of meaningful connection) soon reaches its *limits*. (…) In psychopathology psychic phenomena appear suddenly as something entirely new, in a way which we cannot understand at all. One psychic event follows another quite incomprehensibly, it seems to follow arbitrarily rather than emerge. (Jaspers, 1913, p.27)

> If we try to get some closer understanding of these primary experiences of delusion, we soon find that we cannot really appreciate these quite alien modes of experience. They remain largely incomprehensible, unreal, and beyond our understanding. (Jaspers, 1913, p.98)

However, in my opinion, the concept of understanding is weak because of its subjective nature. It could be counter argued that Jaspers would be unmoved by this commentary, that is, that understanding has some sort of weakness, due to the fact that Jaspers himself would definitely believe that the sort of empathic understanding could be achieved trivially in the first scenario. About the second objection, Jaspers would probably say that, his account is "essentially subjective" because the phenomena in which he is interested are subjective in nature.

Primary delusions, which are ununderstandable, are elementary, immediate and final experiences, precisely opposed to erroneous judgements arrived at via thinking: "The elementary experiences psychologically cannot be influenced opposite to that which is mediated by thoughts. They are primarily without content" (Jaspers, 1913, p.110). This type of delusional state occurs *before* the erroneous belief into which it can crystallise. The only thing the practitioner can do, given that this experience is ununderstandable, is to *interpret* the utterances of the patient, failing to grasp true empathic understanding in the phenomenological sense. So, Jaspers' definition is as much negative as it is positive in the sense of building up the concept. Jaspers takes us all through the definition of the concept of "understanding" to take us to the further concept of "empathic understanding" in order to say that what constitutes delusion (proper) is precisely what lies beyond its limits. Even more, what is said by the patients cannot be truly represented: "There always remains a big leftover of something incomprehensible, intuitively not being representable, not understandable" (Jaspers, 1913, p.82).

According to Kraus (2014), the two definitions of delusion set out above are not incompatible if one thinks they are related from the point of view of experience: delusion is not about our common, shared reality, but about a different reality experienced by the patient. This specificity is not taken in account in diagnostic manuals such as the DSM, however. Then, the altered consciousness of the patient must only lead to false judgements, which are strange and bizarre seen from the outside, but in fact they are present in order to preserve the patient's mental life. Kraus says that even in delusional perception, the delusional interpretation is already given *a priori*. In other words, according to Kraus' interpretation, Jaspers suggests that secondary delusional thoughts are erroneous, but not necessarily incomprehensible, because they come from thoughts formed from an extraordinary experience. However, this is not the case for primary delusions, which are formed first.

I would like to highlight the relationship between Jaspers' notion of understanding being the mark of the mental and delusions being ununderstandable. To Jaspers, ununderstandable psychic phenomena, such as primary delusions, cannot be explained from within; one can only explain their causal relations seen "from the outside" (Jaspers, 1913, p.28). In other words, to Jaspers we cannot have empathic understanding towards the patients, we can only witness from the outside, with perplexity, what is going on in delusions. Even more, if they are ununderstandable, and understandability distinguishes mental states from non-mental states the following question arises: Are delusions a mental state? Jaspers did not directly answer this question but it is implicitly part of the motivation for more recent philosophical approaches to delusions which seek to shed light simultaneously on the resistance to understanding of delusions, with them nevertheless, being mental states, although of an unusual kind.

VII. Delusions as beyond the understandable

Having defined delusions, Jaspers comes to the conclusion that they are an interpretation of the world under a radical transformation:

> We cannot say that the person's whole world has changed, because to a very large extent he can conduct himself like a healthy person in thinking and behaving. But his world has changed to the extent that a changed knowledge of reality so rules it and pervades it that any correction would mean a collapse of Being itself, in so far as it is for him his actual awareness of existence. (Jaspers,1997, p. 105)

Related to this explanation, the impossibility of delusion comes from a different kind of "knowledge of reality". As many contemporary philosophers of delusion assert (especially, Sass (1994a)), Jaspers states that delusional patients seem to take things in a certain, ironic way, different to the mentality of normal people. For patients, their experience "does not always carry the same meaning as that of normal reality" (Kraus, 2014, p.119).

What shall we assume then about Hoerl's interpretation of what is understandable in delusions, if delusions are, precisely, that which cannot be understood? Given that, for

Hoerl, understanding is more than a merely rational explanation of the sort Dennett and Davidson stress, and is more like the preconditions for being rational, one has to conclude that such preconditions are what is consistently failing in delusion.

Therefore, Hoerl's account of what is understandable helps us to interpret what Jaspers meant or might have meant in contemporary terms. However, it is a different question whether the idea of a failure of the preconditions of being rational can correctly and positively describe delusions. If one links understanding with examining the conditions of possibility of rationality and one holds, with Jaspers, that delusions are ununderstandable, that is not enough to shed light on the nature of the failure of rationality. For example, must it be, paradoxically, a *mental* failure of rationality? In other words, it is not clear that even Hoerl's attempt to shed light on what Jaspers might have meant by understanding actually helps shed light on delusion. This question will return in a later chapter on Sass and Campbell.

VIII. Conclusion

Jaspers' discussion of delusion continues to have three main areas of influence.

Firstly, he has set the agenda for future definitions of and frameworks for thinking about delusion within psychiatry and philosophy. Current definitions might contest Jaspers at some point, but still most of his definition prevails, especially if it is part of the DSM definition. However, this influence is deeper; it has determined what should be relevant since those times. Although reductionist approaches are still important in contemporary psychiatry and philosophy, as they were in Jaspers' times, the pressure to characterise delusions in person-level mental terms remains.

A second outcome consists in the wave of phenomenological approaches that are active in psychiatry, inspired by Jaspers, Husserl and Heidegger. Influential contemporary philosophers such as Ratcliffe, Sass, Gallagher, etc. who question the current analytic traditions, find their roots in the latter phenomenological theories.

For this thesis, the third influence is the most important. Jaspers' preferred approach to delusions suggests an apparent contradiction, which is that *the general* concept of delusion is understandable only by being non-understandable *in particular instances* of the phenomenon. In other words: Jaspers thinks he can grasp the phenomenon of delusion (he can *understand* delusion) by proposing that one cannot make sense of the

patient's utterances (therefore, delusions are a particular instance of non-sense). However, the matter of delusions does not end with Jaspers' analysis. One of the reasons is that there seems to be a tension in Jaspers in the sense that "understanding" is at the heart of his psychiatry in all of its concepts, *except for delusions*. This tension is bigger if one thinks that being understandable is a mark of all the other mental states. Does that mean that delusions should not be considered as mental states? This does not seem plausible.

As I will discuss in subsequent chapters, philosophers of delusion have hoped that delusions might be understandable *as something else*, some more familiar mental state which can explain but circumvent Jaspers' pessimism, since his concept of ununderstandability seems to capture the fact that delusions are actually strange. So, the question still remains: what sense can we make of delusions?

Before later turning to more recent phenomenological work and other philosophical interpretations I will begin in the next two chapters with the debate in the philosophy of delusion between doxastic and antidoxastic positions.

CHAPTER II- ANTIDOXASTIC ACCOUNTS OF DELUSION

I. Introduction

A common assumption about delusions, usually endorsed by clinicians and other specialists, is that they are erroneous beliefs. Although some authors questioned these assumptions during the 19th century (Cfr. Berrios, 1991); and Jaspers (1913) thought that seeing delusions as beliefs was an oversimplification of the phenomenon, it was not until the 1990's that the doxastic status of delusion was questioned in a significant way again. Since then, several different accounts of the nature of the phenomenon have been proposed. In the literature on delusions, it is a matter of debate whether delusions constitute a belief as such, as opposed to being some other state of mind such as an empty utterance, a product of imagination, or an expression of some other kind of mental state. Theories that consider delusions as beliefs are doxastic, while the ones that consider them thoughts of some other nature are antidoxastic (or non-doxastic).

In this chapter, I analyse some representative antidoxastic accounts and set out what the new assumptions of this approach are, along with some critical analyses of the proposed stances. My main focus will be to highlight the theoretical claims that are implicitly and explicitly established through the thesis that delusions are not beliefs. My intention is not to assess the validity of the theories as such, nor to judge whether their claims are an adequate description of delusions.

My aim is to emphasise the aspects that the author states play a certain role in explaining the nature of delusions and to put these aspects together explicitly. In the end, this kind of analysis, performed on different theories and authors, will hopefully give us some understanding of an important part of what constitutes the current theoretical framework of delusions and their underlying assumptions. However, as I propose at the end of each section, the representative accounts analysed are not able to overcome some major difficulties in their current form.

II. Delusions as imaginings

The first anti-doxastic approach I will consider is Gregory Currie's theory that delusions are pathological imaginings. This idea is based originally on Chris Frith's (1992) theory

that the various symptoms of schizophrenia are generated by a disorder in the capacity of identifying imaginings at a subpersonal level. This disorder comes, according to Frith, from an impairment of efference copying, which results in an impairment of action monitoring, which is comparable to impairment in intention monitoring. According to Frith's influential theory, an 'efference copy' is made when the brain issues an order, for example, to move a leg, and this order generates necessary information to the relevant parts of the brain and body to carry out the action and also generates a copy of the order to be filed in the brain. It is this that enables the brain to distinguish subsequently between actions it has initiated and other externally influenced movements of the body. The latter piece of information is called the efference copy. Therefore, if such an efference copy is missing, the brain can initiate an action without knowing it was issued by itself. This is, according to Frith, an important form of impairment in the self-monitoring mechanism which explains some of the symptoms of schizophrenia. It prompts Currie to suggest: "Let us consider the idea that the schizophrenic patient is someone who has lost the distinction between what he or she imagines, and what he or she experiences" (Currie, 2000, p.174).

Currie argues that delusions are not beliefs, although they look like them in some aspects. To Currie, delusions are imaginings which are mistaken by their subject for beliefs. Thus they count as "*cognitive hallucinations* which occur when a mental state of one kind (an imagining) presents itself to the subject as a mental state of another kind (a belief)" (Currie, 2000, p.176). Hallucinations, in general, standardly occur when a non-veridical experience is taken as veridical:

> And hallucinations sometimes give rise to delusional beliefs; I have an experience as of P, and I come to believe that P. Let us broaden the notion of hallucination somewhat as to include what we might call "cognitive hallucinations", which occur when one mental state of one kind presents itself to the subject as a mental state of another kind. (Currie, 2000, p.176)

To Currie, this means that the actual hallucination *is that one believes that P* given a certain (non-veridical) experience.

Although Currie's theory contrasts delusions, understood as a form of non-doxastic state with beliefs he has, at the same time to explain why it is tempting to mistake delusions

for beliefs. His answer is that imaginings are simulative states, that is, states that "mirror" some of the features of other mental states, for example, as when adults take part in reading fictional stories, or when children engage in role-playing. Such simulative states are hard to distinguish from beliefs in their inferential roles, because both beliefs and imaginings will lead to a similar system of inferences when recalled, which could be quite complex. The complexity in the inferential structure of imaginings to which Currie is referring can be found in fantastical literature, where characters act in ways that are logical given their assigned roles. For example, one could infer the appropriate complex behaviours of a fairy or a famous magician, although none of them actually ever existed. This suggests a key tension in Currie's account. On the one hand, simulative states have some of the same inferential properties of belief states and so it seems understandable that they could be confused. On the other, if delusions are a simulative state that rules them out from being a belief state.

Another characteristic of delusion as an imagining is that the subject does not always act according to it. Currie explains in his account that the subject may lose, intermittently, his capacity to monitor his own active thinking. This is a fundamental step in Currie's account. The mechanism by which delusions are formed is the following: First the subject imagines something. After that, the subject draws *inferences* from this *imagining*, while it still retains the status of imagining. However, the subject can suddenly lose his capacity of detecting (monitoring) that this thought is only an imagining. Because of this lost capacity, the content of the imagining comes out as something that the subject has simply thought. Therefore, the hallucination takes place, consisting on the fact that an imagining is presented to the subject as a belief. This then leads to other inferences, but as the inferential consequences of a belief that the subject has formed in the normal way, rather than the original imagining's inferences.

On his account Currie aims to explain the phenomenon of thought insertion. Thought insertion is the delusion in which the subject maintains that an external force (for example, another person that he knows, or a celebrity) inserts thoughts in the patient's mind. Therefore, the patient maintains that he is not the author of some of his thoughts, and this usually causes him distress and anxiety. Currie explains that the failure of monitoring his own thoughts, due to the lack of an efference copy, may cause the sense of thought insertion in patients. Although it is the case that efference copies were proposed originally for monitoring actions by Frith, Currie takes it further to take in

account thoughts as well. Therefore, a point in favour of Currie's account is that it has a lot of explanatory power, which is very compelling.

Currie argues that imaginings have unique characteristics that help to explain the similarities between delusions and imaginings. To do this, he suggests that there are two different kinds of imaginings: factive imaginings, and perceptual imaginings. Factive imaginings are instances in which the subject theorizes what would happen if such and such were the case. Perceptual imaginings have features in common with the corresponding kinds of perception and consist of imagining stimuli collected by the senses, such as auditory and visual imagery. Empirical evidence suggests that perceptual imaginings can be confused by normal subjects with genuine perceptions. To Currie, the fact that some schizophrenic patients have visual and auditory hallucinations demonstrates that these patients could have formed a perceptual imagining and then mistaken it as an external hallucination due to their lack of imagining-forming awareness. If Currie is right that delusions in general are *factive* imaginings, subjects experience them with *subjective conviction*, akin to belief. Thus Currie says that this is not, after all, an exclusive characteristic of beliefs.

> What is emerging here is a picture of (some) delusions as states which, on the one hand, are imaginings but on the other, have a feature that most imaginings lack: a feeling of subjective conviction on the part of the subject. One might object that this feeling of subjective conviction is exactly what distinguishes beliefs from imaginings; if that were true, our claim would be as problematic as the claim that there is a peculiar class of bachelors—the married ones. However, we deny that beliefs are identified by a feeling of subjective conviction. (Currie, 2001, p. 162)

Currie's contribution to the non-doxastic accounts is that imaginings can substitute a belief-like state in the case of delusion. In favour of his account, Currie expresses the following characteristics of imaginings:

- Imaginings are capable of having many of the inferential characteristics of beliefs.
- Imaginings are more easily triggered by perception than beliefs.
- Imaginings can have wild hypotheses in response to an odd experience.

- Imaginings are not apt to be revised in the light of evidence.

Due to all of these characteristics, Currie states that it is impossible to empathize with a subject who has lost the capacity of distinguishing his imaginings from his beliefs:

> Those of us who have not experienced these conditions are not able to empathize with people who suffer from unilateral neglect, who have lost the ability to recognize people by their faces, or who can no longer negotiate their way through space. Delusions, we suggest, resist empathy in just the way these disorders do. (Currie, 2001, p.162)

Currie seems to suggest this because the lack of empathy would make delusions similar to other brain damages which are hard to understand, such as prosopagnosia, unilateral neglect, and other sub-personal level damages.

One of the other important contributions of Currie's view is precisely that he treats delusions as something familiar to everybody, that is, an imagining. By doing this, he sheds light on something apparently mysterious – the nature of delusions – by comparing them to something familiar: acts of imagination. I will return both at the end of this chapter and in the final chapter to the question of whether this can yield a satisfactory account of delusions.

III. Objections to Currie

Objections to Currie's account correspond to two main enquiries: Are the functional roles of imaginings sufficiently explanatory to account for the oddness of delusion? And, why can't imagining be doxastic in nature? I will dedicate the first three counterarguments to the first question, whilst the last one will be about the latter one.

In the first place, are imaginings the correct ontology even though they rarely prompt to action? Currie acknowledges that his theory cannot explain why sometimes delusions are acted upon: "perhaps at this point their thoughts have arrived at the status of beliefs, though they will often not retain this status indefinitely" (Currie, 2000, p.177).

In addition, Currie is not prepared to give a description of what roles an imagining should have in order to be a substitute for a belief. The belief role includes the fact that they are endorsed, sometimes acted upon, and have a strong emotional charge. That is,

imaginings do not have the functional role of prompting to action and generating strong feelings, as beliefs do.

Currie's answer would probably be that the subject has no awareness of the delusional state being an imagining, because the delusion necessarily becomes a second-order state, which in turn generates a belief. However, this would be giving to imagination a new, different role that it is not quite imagining because, as, Currie himself says, it would have to become a belief before it can actually prompt into action. So, in the better case, his account is impossible to prove, because how could we know that a certain thought was an imagining (or any other sort of thought) before becoming a belief? Although one has to concede that Currie's account is a plausible mechanism that works behind the belief that has gone wrong, in the worst case scenario, it is an ad hoc explanation which describes what delusion does, in which the role of imagination becomes unimportant.

A further question that is relevant to Currie's philosophical approach is, if delusions are merely disorders of imagination, why are these disorders relatively circumscribed, since imagination is certainly not circumscribed at all? For example, why it is not the case that the patients imagine, instead, that they are rich and that they have a very satisfactory life, instead of imagining threatening delusions? Related to this, we have the fact that people with delusions can still imagine other scenarios and other interactions with people without becoming delusional about it. Schizophrenia would be too simple to explain if any imagining became a delusion, but this is not the case. So we have another explanatory gap that Currie needs to close in order to have a satisfactory account.

Yet another objection against Currie is that he does not give an independent anti-doxastic account against doxasticism: he explains how it would be possible to think that a certain imagining has become a delusion, but he does not explain why delusions cannot be doxastic in the first place. This is only implicit in the underlying fact that delusions show contradictory features, while normal beliefs tend to be more consistent, but as doxastic arguments show, that is not sufficient to tell why or how normal beliefs are different from delusions. As I have already hinted, this is a feature of the form of argument Currie deploys, to which I will return at the end of the thesis.

IV. The concept of bimaginings

The second anti-doxastic account of delusions that I will discuss can be seen as a development of Currie's claim that delusions are imaginings. In his (2008) paper "Imagination, Delusion, and Self-Deception", Andy Egan suggests that delusions could be an intermediate, sui generis state called "bimaginings" which is in some aspects like a belief and in some aspects like an imagining.

This philosophical approach has been motivated by two apparently contradictory aspects of delusions. The first one is that delusions show many characteristics of being belief-like, centrally that patients show strong endorsement towards their delusional idea. However, the other aspect is that they are still often inert in many behavioural, affective, and inferential aspects. This contradiction is a key puzzling feature of delusion. Egan's idea is to combine two different states - belief and imagining - into a gradation that could collect all the characteristics that delusions show and mix them together to achieve a new, different, comprehensive view. Egan's definition of this new state is the following:

> I want to suggest that, instead, we ought to say that delusional subjects don't straightforwardly believe the content of their delusions, nor do they straightforwardly imagine them. Instead, they bear some intermediate attitude- we might call it "bimagination"- with some distinctive features of believing, and some distinctive features of imagining. (Egan, 2008, p.263)

A key question is whether combining imaginings and beliefs can actually carry the weight of delusional thinking. Since imaginings and beliefs are two known entities, how are these related to a seemingly different state? Given their similarities and their differences to each other, will joining them give us the desired result of accounting understandably for the features of delusion? This specific question has not been addressed by Egan, and so we shall discuss it here in order to evaluate the contribution he actually makes to the non-doxastic views.

A preliminary question is how much delusions are like beliefs in the first place. This is, perhaps, closer to our intuition because delusions share many of the main aspects that may form the criteria of belief. For example, subjects, when directly questioned, say that

they believe in their delusion[4], and they sometimes behave in the "appropriate" way concerning their delusion. Also, although it is argued that beliefs do not always lead to action, and that sometimes they are compartmentalized, this is sometimes the case with normal beliefs. In compartmentalization "limited resources in memory and attention force subjects to store beliefs in different compartments of their mind, and prevent subjects from attending to the conflicting beliefs all at once" (Bortolotti, 2010, p.88). There are many doxastic accounts that will give very good arguments for why delusions are sufficiently like beliefs to count as beliefs (Bayne and Pacherie 2005; Bortolotti 2010; Stone and Young 1997).

What is less clear is in which aspects delusions are not like beliefs. Egan maintains that beliefs should respond to evidence in the right kind of ways, and delusional subjects seem to ignore the relevant evidence and just to take the evidence that will support their actual delusional view (self-serving bias). Egan also argues that delusions seem to be more circumscribed than other beliefs because they are not always inferentially connected, and they do not always promote behaviour. In addition, sometimes delusions do not display the correct emotions in an affective form.

Is this a plausible argument for a non-doxastic account? As I have just highlighted, one counterargument from doxastism is that normal, non-delusional beliefs can also be non-inferential because sometimes they can be compartmentalized. When beliefs are compartmentalized it means that those beliefs will be independent of other sets of beliefs, and usually, in contradiction with those other sets of beliefs.

Later in his paper, in attempting to make space for the very notion of bimagining Egan argues that normal belief does not always show all the characteristics that one would normally attribute to it (Cfr. Egan, 2008, p.271). In his example, when the lights fail in a house, a person watching TV could say, "Oh, no, I can't watch TV anymore; I should go and check my email", forgetting momentarily that the electricity supply would still be down, therefore having a belief, which was non behaviour guiding. Whilst this is used to justify the possibility of bimagining, however, this very same example can be used to defend the doxastic thesis, by saying that normal beliefs are compartmentalized and non-inferential, and that this is a common characteristic of normal beliefs. Therefore, Egan's point that beliefs are not always action-guiding is made. But it is not the case that

[4] Examples of these assertions can be read on "Pathologies of Belief" (Davies and Coltheart, 2000)

bimaginings are the only possible answer, because we have seen that normal beliefs could show some "odd" characteristics as well.

In addition, it has been shown that normal subjects in normal circumstances also suffer from a self-serving bias, where people tend to have new beliefs that will conform to the previous background of beliefs they already had, especially when the actual evidence is ambiguous. It is also the case that beliefs do not always promote behaviour. For example, studies show that normal people suffer from poor self-prediction (Bortolotti, 2010, p. 174): when students were asked if they would cheat on exams given the chance they said they would not, but in the actual case the percentage of the ones who did was higher than predicted. Hypocrisy is also a case where normal subjects fail to engage in behaviour that accords with their beliefs (Aronson, 1999). Also, delusional patients sometimes display their emotions in an appropriate way: subjects get very stressed about their delusional idea, and they sometimes engage in aggressive behaviour more in line with their claims. So, Egan has not given so far a conclusive argument against doxastic views.

Egan could reply that delusions are doxastic only in the sense that they have some actual characteristics of beliefs, but that one cannot ignore the characteristics where they are closer to imaginings as well. They are not like beliefs in all respects. However, it looks as though Egan only chooses the characteristics of belief as he sees fit, moulding them to his convenience so as to argue that the variation from the paradigm of actual beliefs justifies the possibility of bimagining whilst denying that it implies, directly, that delusions could be non-paradigmatic beliefs. He can grant that beliefs can be odd when and only when it is convenient to him, which makes his stance too *ad-hoc,* and still does not undermine doxastic arguments. The problem here is whether beliefs have these characteristics intrinsically, as doxastic views maintain, or if they just show odd characteristics sometimes, as it is in Egan's ad hoc view. Both views seem equally plausible so far, depending on the role one wants beliefs to perform.

I will consider now if the idea of delusions being like imaginings is more adequate. Delusions are like imaginings, in Egan's view, because they display the right kind of evidence independence, the right kind of inferential circumscription, the right kind of behavioural circumscription, and have a different affective impact than a belief. So, in this sense one wins explanatory power by saying that delusions have many similarities with imaginings. But are the similarities enough to make an imagining have the

functional role that a belief does? Are imaginings sufficiently strong in order to substitute the role of a belief? Egan's idea is that delusions are just states which are in the midway between beliefs and imaginings, therefore never being one or the other completely.

How would a bimagining work in order to sometimes be like a belief, and when would it become just an imagining? One fundamental flaw of Egan's theory is, precisely, that patients sometimes present a very mixed state in delusions, where they actually work with a belief's inferential power but do not engage in the appropriate behaviour, or any other combination. These cases are not considered in Egan's account, but are usually common in subjects with delusions. In this case, Egan could say that these states could also exist, but that they are very pathological, as he expressed in his paper "We shouldn't expect to see the peculiar hybrid roles all over the place because they're pretty maladaptive" (Egan, 2008, p.275).

There are two key objections to Egan's account. The first objection questions whether bimaginings, as a combined state, are sufficiently explanatory of the phenomenon. The second objection is whether the given explanation is truly enlightening about what a delusion is. I will refer to those specific questions in the following paragraphs.

Regarding the fact whether functional roles, as Egan proposes, are explanatory, one can analyse this first case. A fundamental flaw in bimaginings is that imaginings are too weak to explain how patients sometimes engage in behaviour. Therefore, imagination's only role could be to act as a promoter of the behaviour correspondent to a belief, but not to produce behaviour directly. For example, one could imagine frequently that one wins the lottery at some point. This exercise in imagination could promote one person to buy a ticket, but only when the imagining is replaced by the belief that one could actually win, is the action done. That is the essence of having an imagining and not a belief. Egan's answer could be that in those cases we would have to engage with the "belief" part of the bimagining, because imaginations are not usually acted upon. However, if it is more than an imagining, why do some delusional subjects constantly change their mind about the state of their idea? It would be improbable that the idea changes its status intermittently because once a belief is formed it usually does not go back to being an imagining. A relevant clinical case where a patient with Capgras changes his mind intermittently about the identity of his wife can be found in Lucchelli and Spinnler (2007). Therefore, the functional role proposed cannot work empirically as it is, because

Egan intends his account to work in all cases of delusion, including specific cases as delusions in the Capgras syndrome.

Another objection to the functional role of bimagining is that Egan's argument maintains that the belief role is so broad that it could make some other intermediate states plausible, not just bimaginings. However, none of these proposed intermediate states have been identified yet.

One thing that puts more tension in Egan's account is that he suggests that states like bimaginings are pathological:

> We shouldn't expect to see the peculiar hybrid roles all over the place because they are pretty maladaptative (…) So there's good reason to expect to see the kind of clustering that we do – deviations from it are signs that something has gone wrong. (Egan, 2008, p.275)

But to state that is very tricky, perhaps a contradictory idea to Egan's own philosophy. To start with, we cannot be very sure what it means to have a state that is defined by a gradation of two other known states because of the following: Having a state that combines characteristics of two natural kinds is not the same as having a state that is between both, as in a gradation. In the first instance we could simply list the various characteristics, but in Egan's method we would need to measure where in the spectrum a specific delusion is placed. Egan has not suggested a method for doing this. So, we do not know how to "measure" this gradation in the first place. However, supposing that we understand this position, that delusions are part of a gradation, it cannot be stated whether the delusion is in a state other than somewhere in between imaginations and beliefs. This notion is problematic because it is in tension with Egan's idea of these being maladaptive states, because in states which are in between, one cannot have a definition of maladaptive: one can only have the concept of being closer to one state or the other, but since none of them are intrinsically more adaptive, the distinction of its adaptiveness loses its sense. The question of being pathological, given a gradation, only makes sense put in context, because there are not intrinsically right or wrong states, there is only a combination of two states that fulfil different functional roles, and therefore, having different functional roles, on its own, cannot be pathological. Being maladaptive would also imply that there are some specific finite functional roles

because pathological states are defined by certain characteristic features. However, this is not compatible with Egan's notion of a continuum between imaginings and belief states. But Egan still has a way of defending his own position by defining a new role for beliefs, which can be compatible with his idea of a continuum. However, it is not clear that Egan actually wants to redefine the idea of belief.

There are fundamental reasons why I think that Egan's view of delusions as bimaginings is ad hoc. It is not so much about the fact that he tailored his theory in order to explain delusions with two familiar functional roles; rather, it has to do with the tension that his account generates within its own postulates. My worry regarding bimaginings is about the balance between mixing two familiar functional roles and the idea that they are infrequent. Egan himself advances the question: if bimaginings are possible, then why we do not see more instances of them? It is precisely his own argument of why there are not more of them that becomes an argument of why there are not any at all in the first place. The tension in his account is between granting bimaginings enough flexibility to have mixed propositional functional roles of two propositional attitudes, and, the fact that you cannot have many of them. For example, bimaginings cannot fit easily into rational interpretation. In other words, the claim is that Egan's account is ad hoc because there is a tension in him a) offering a mix of functional roles and b) explaining the rarity of delusions by saying that these roles are in tension. Pointing out the clash between the functional roles of imaginings and beliefs to deluded subjects should force thinkers to give them up. But they do not give up their delusions.

The final criticism of Egan's view is that it is not clear to what extent his idea is new and original. The idea of there being a sui generis state is valuable in the sense that it would be precisely a new state whose characteristics are yet to be defined. However, that the exact combination of an imagining and a belief would result in the desired sui generis state is doubtful, mostly because Egan's argument against the doxasticism of delusions is still too weak, and the arguments for delusions being imaginings are better said in other accounts, such as Currie's. Also, there is the question of whether a sui generis state would be useful in itself. In a first approximation, it could be contended that the concept of bimagining is informative due to the fact of it being part imagining and part belief. However, as I argued earlier, the gradation between a belief and an imagining is

not well defined enough to have a clear idea of how the dynamic of the combined functional roles has to be applied.

One would have to consider the fact that sui-generis would bring the intuition that delusions are non-understandable, because they would be a state which cannot be recreated phenomenologically by the mind in any state other than delusional: a state that cannot be recreated in other people's minds other than being delusional could probably not be empathically understood. It could be that empathizing with delusions is like Daniel Dennett's problem of the lemonade stand (Stich, 1981). In this problem a boy who sells lemonade gives someone the incorrect amount of change even though he knows perfectly how to do this simple transaction. The failure in both cases (the lemonade stand boy and the delusional case) is inexplicable, but somehow, in the case of the seller, it is easier to relate to the failure. In the case of delusions, one should take into account that the thought of being non-understandable would be in contradiction to the insight that many patients show while being delusional, in which they are able to give reasons and to sustain their point of view on many occasions, while stating that they do not feel any different from normal.

V. Delusions as second-order states: The Delusional Stance Thesis

The third non-doxastic view I will discuss is that of Graham and Stephens. Their account proposes that delusions are second-order attitudes or thoughts. They call their approach the Delusional Stance Thesis (DST). It is expressed in two parts:

> DST-1: All delusions are higher order attitudes that constitute a kind of stance taken towards lower order mental contents or intentional states which may be of a variety of different sorts (including, but not restricted to, beliefs). (Graham and Stephens, 2007, p.194)

> DST-2: The higher order or delusional stance involves the deluded person's identifying himself or herself with the relevant first-order contents, so much so that he or she persistently maintains those contents, incorrigibly, in the face of strong counter-considerations and with a disturbing lack of diagnostic or first-person insight. (Graham and Stephens, 2007, pp.194-195)

Graham and Stephens assert that the pathology of delusions is due to their second-order character rather than anything to do with the nature of the first-order thought on which they are based:

> We shall argue, similarly, that the deluded person's problem does not lie in the content or nature of his/her first-order intentional states, but in his/her failure to appreciate the role that these states play in his/her psychological economy. It is not *what* it is thought, but *how* it is thought that explains the pathological character of delusions. (Graham and Stephens, 2007, p.195)

Graham and Stephens assume the following characteristics of beliefs in order to dismiss them in delusions later on:

- Beliefs have content.
- Beliefs are held with conviction.
- Beliefs guide reasoning and action.
- Beliefs will have appropriate effects on the emotions.

They suggest that these characteristics are not sufficiently present in delusions drawing on discussions in other philosophical accounts, such as Sass's (1994), Berrios' (1995) and Currie's (2000). Therefore, they are in agreement with other antidoxastic accounts.
The way the Delusional Stance Thesis works is as follows: the patient has the mistaken belief that he/she has the belief P, P being the content of the delusion.

> She [the subject] answers "Yes" to the question "Do you believe that the sky is falling?" because she second-order believes that she (first-order) believes that the sky is falling". (Graham and Stephens, 2007, p.201)

The first-order attitude does not need to be a belief. The deluded subject might be just considering, imagining, thinking, or believing the first-order thought. The second-order attitude will be what determines that the subject will be convinced of the truth that P, but she will signal in many ways that she does not have the belief that P. That is, the subject will be able to think "I believe that P", but when the relevant belief P has to be put in action, the subject will just have an inert behaviour towards it, unless she is

engaging in a second-order thought about it. Summarizing, the first-order attitude will range over different propositional attitudes whilst the second-order attitude would be a stance towards those first-order attitudes. Graham and Stephens conclude: "Thus our account accommodates the clinical observation that patients to whom it seems appropriate to attribute the delusion that p in some instances exhibit and in other instances fail to exhibit the behaviour characteristic of someone who believes that p" (Graham and Stephens, 2007, p.203).

They argue that the delusional stance approach to delusions explains the following characteristics of the phenomenon:

- Self-identification: Subjects having the delusion that p identify themselves with the act of thinking that p (although in the Delusional stance the latter is not necessarily the case for the subject).
- Resistance: If the subject has the delusion that p, she will resist the idea of changing her mind about the first-order thought that p.
- Lack of insight: Patients do not acknowledge that there is something wrong within them rather than with the world because the second-order state allows them to believe that they believe something (Cfr. Graham and Stephens, 2007, p.208).

VI. Criticism of the Delusional Stance Thesis

Criticisms of Graham and Stephens' account consist of the following three general objections: Is a second-order thought empirically adequate to describe delusions? Is this stance sufficiently different from doxastic accounts? How is the fact that delusions are pathological explained? And, is this stance being too ad hoc? I will go through them in detail in the following paragraphs.

As I mentioned in the first paragraph, the preliminary question to Graham and Stephens' characteristics of delusion is: Are these not characteristics of normal (doxastic) beliefs? If so, then the account is not sufficient to account for delusions. Is it not the case that once a belief is generated it shows the three characteristics mentioned (Self-identification, Resistance, and Lack of insight)? Because normal subjects usually want to maintain their own ideas whether or not there might be contradicting evidence, these three features of belief are not uncommon. People self-identify with their beliefs: when

normal subjects are asked about their beliefs they can often endorse them with reasons. Self-identification may be a necessary characteristic of delusion in order to distinguish it from other pathologies such as OCD. Still, it is not enough reason to motivate a non-doxastic view of delusions because both second-order beliefs and first-order beliefs have the self-identification characteristic.

The second characteristic (resistance), proposed by Graham and Stephens, is undermined by doxastic accounts, such as Bortolotti's (2010), that argue that it is normal for people to change their beliefs once they have a definite idea: people tend to keep their previous beliefs in cases of ambiguous, inconsistent or dubious evidence. Some people also show some lack of insight by being stubborn about their beliefs, as in the case of unrevisable beliefs, such as racism or religious ideas. So, although they help the overall explanation, they are not exclusive to the DST. However, one might argue that resistance is a defining factor because first-order beliefs are more resistant to change than second-order beliefs in most cases. That is because second-order beliefs are dependent on these first-order states: my belief that there is a cat on the mat depends on there being a cat on the mat. However, delusion is the one anomalous case where second-order beliefs work in exactly the opposite way to how they would normally work. One could argue with a delusional patient that the evidence points towards his wife not being a robot, but he would argue that his subjective (second-order) thought, that is, that he believes so, is more reliable than any empirical justification. Therefore, in any case, if delusion were to be a second-order belief it would be an anomalous one.

I think that their claim that second-order attitudes generate a lack of insight in the patient is one of the strongest points of Graham and Stephens' account. However, it still has the characteristic of being ad hoc. They have not provided a rationale for why the delusional stance typically lacks insight.

There is still the question of how the DST uniquely explains that delusions are pathological. The delusional stance seems to be very similar to cases of common hypocrisy, where the person thinks that she is congruent with her thought and actions, but she is not. It is similar to cases where one person says that she is a very religious person, and when questioned if she is a good person she will answer "Yes". But outside the church the same person could be cruel and dishonest, and still believe that she is a good, religious person. This phenomenon is called double book-keeping, and is not uncommon amongst the population. In this sense, the delusional stance is not unique

to delusions, and cannot explain why delusions are pathological, since a common phenomenon, like hypocrisy, can be explained in the same terms as those of the delusional stance.

One possible answer is that hypocrisy could be pathological, that is, delusional, when engaged in double book-keeping, which is abnormal in certain ways. However, there are other instances of thoughts, such as religious thoughts, that are held with persistence and are not delusional. So, resistance, on its own, cannot be the mark of pathology in delusions.

Overall, Graham and Stephens' account is too ad hoc, in the sense that its explanatory power comes from describing what a delusion does. This is similar to a sui generis account, except for the part where they define the delusion as a second-order state. However, underpinning delusion to those second-order states does not help to explain what is actually pathological in delusions. Graham and Stephens deploy their account to explain delusions as something else quite familiar to normal subjects. However, such explanation does not cover the fact that delusions are pathological. The transition between a second-order thought and a pathological thought is never explained. It is still a mystery how something familiar becomes pathological given their account.

VII. Keith Frankish's Two-level Framework

My final example of a non-doxastic approach to delusion is Keith Frankish's account. Frankish defends what he calls a two-level framework of beliefs. The reader might consider that it is odd to have an account of "beliefs" in a chapter about antidoxasticism. Frankish's account is a stance that encompasses all types of beliefs, and only a particular subsection of this framework is designed to account for delusions. For that reason the stance proposes a non-doxastic account, but overall, his account is a general framework of beliefs.

Also, Frankish's work disagrees with a doxastic stance in which all beliefs show some degree of irrationality. In his (2012) paper, Frankish explains his objections to Bortolotti's (2010) claims about the doxastic nature of delusions. The key element here is whether delusions actually constitute a belief, despite the fact that other options, such as mere acceptances, are available. In particular, Frankish argues that delusions are some sort of

mental state that he calls policies which are epistemic in a specific way. Therefore, his account is both against a doxastic account and proposes an antidoxastic alternative.

In general, arguments against the claim that delusions are a species of belief come from the apparent difficulty of fitting delusions into the different notions of what constitutes a belief. The main objections against delusion as a doxastic state usually consist in the intuition that they violate rationality norms, and that delusions do not work in the same way non-delusional beliefs do. In particular, Bortolotti and Frankish focus on the rationality criteria explained in the next paragraph in order to discuss the doxastic nature of delusions.

Here I briefly summarize Bortolotti's (2010) stance, although it is described in detail in the next chapter, dedicated to doxastic accounts. According to her, most of the relevant features of beliefs can be classified under the following three criteria of rationality:

- Procedural rationality – beliefs ought to be well integrated with the subject's other beliefs and intentional states.
- Epistemic rationality- beliefs ought to be well-supported by and be responsive to the evidence available to the subject.
- Agential rationality- the subject ought to be disposed to provide reasons for his beliefs and act on them in the relevant circumstances (Cfr. Bortolotti, 2010).

In her (2010) book Bortolotti concludes that, although delusions sometimes fail to address these norms of rationality, there are some cases in which they actually adhere to them. To make this case even more complicated, she also argues that non-delusional normal beliefs also frequently fail these three criteria of rationality in normal, everyday environments. She concludes that delusions are beliefs, and they belong to a continuum of beliefs with normal beliefs.

In 2009, previous to Bortolotti's book, Frankish published his paper "Delusions: a two-level framework" (Frankish, 2009), where he explained his conclusions about the nature of delusions. After the publication of Bortolotti's book, Frankish wrote his (2012) paper in which he criticises Bortolotti's continuum of beliefs. Having these considerations in mind, I will proceed to analyse Frankish's theory of delusions in detail.

VIII. The Two-level framework

According to Frankish, the word 'belief' is used in a number of different ways in folk psychology, philosophy, and scientific psychology. His theory tries to capture some of the different properties exhibited by beliefs and belief-like states. One of the virtues of the theory advanced by Frankish is, precisely, that it tries to capture many of the different properties exhibited by beliefs and belief-like states in a framework constituted of two levels. Each level has different agential, epistemic and procedural roles. It is worth noting that this framework is designed to encompass all types of attitudes, including delusional and everyday ones. It is useful to note that, although Frankish's account is mainly about levels of beliefs, he applies his classification to other attitudes, such as desires, acceptances, and policies.

1) Level 1

Frankish uses the phrase 'Level 1 beliefs' to describe pragmatic attitudes. They are also defined as behavioural dispositions. According to Frankish, these are sustained mainly for practical, everyday reasons. One example of these attitudes is the one that a lawyer would adopt in court by saying and acting as if his clients were innocent although he might think otherwise. They have the following characteristics:

- Non-conscious, not introspectable.
- Passively formed.
- Graded.
- Holistic- in Frankish's terms, holism is "simply to be disposed to behave in ways that would be rational on the assumption that p is true, given one's other beliefs and desires" (Frankish, 2012, p. 24).

2) Level 2

Frankish's 'Level 2' beliefs primarily consist in personal commitments. In other words, these are beliefs that have functional roles. These work as an underlying policy that has to be followed unless special circumstances prevent it. Committing to a policy, Frankish explains, is just 'making up one's mind'. They have the following characteristics:

- Conscious.
- Controlled.
- Discrete and binary.
- Context independent.

The fact that these level 2 beliefs are conscious and controlled is not entirely surprising, since there are very few (if any) theories that would postulate that beliefs are not this way. Bortolotti's arguments against the rationality of normal beliefs do not suggest that they are unconscious or uncontrolled; the accusations of irrationality are of the style of being unconsciously biased, but that never implies that the belief is unconscious. Frankish highlights that level 2 beliefs are very much like controlled policies: "We can make such commitments, I argue, because conscious reasoning is to a large extent under our personal control. I defend this view" (Frankish, 2012, p.24). Frankish defines beliefs as policies. This is a key claim which is not explained in detail in Frankish's previous works related to delusions.

As previously described, level 2 beliefs are functionally discrete and binary. By discrete he means that they can be selectively acquired, recalled, or lost, whilst by binary he means that either one believes in something, or not: "For any given proposition, one either has or does not have a level 2 belief in it (that is, one either has or has not endorsed it as a premise)" (Frankish, 2012, p.25). This binary property is contrasted with level 1 beliefs, in which he states that one can have different degrees of confidence unconsciously attached. Since level 1 beliefs are not conscious, he claims that it is not accurate to think about these attitudes as if one believes them or not, one can only say how attached one is to these behavioural states.

Another characteristic of level 2 states is that they exist as a pre-defined policy and, although they can be initially formed for epistemic reasons and, once they are already shaped, they are not easily moved by the available evidence. Therefore, according to Frankish, they are not committed to epistemic rationality: "norms of epistemic rationality have no significant role in constraining attributions of level [2] belief" (Frankish, 2012, p.26).

Also, Frankish explains that level 2 beliefs are committed to procedural and agential rationality only because of the role this kind of rationality plays in decision making, but it has to be understood that decision making is not entirely rational and many factors may influence such decisions. Some of the things that influence decision making are

factors such as motivation and competence, as well as interference of level 1 behavioural dispositions.

According to Frankish, cases of non-doxastic acceptances correspond to level 2 behavioural dispositions. These are something like a working assumption that one is not necessarily committed to (although one might endorse it, in certain cases). This state is clarified by an example:

> Now, not all acceptances are beliefs. One may accept a proposition for pragmatic reasons, such as professional ethics or loyalty to a friend, and the acceptance may be restricted to a certain context. For example, a lawyer may accept that her client is innocent in her professional life, but not in private. We might call this 'professional belief', but it would be misleading to describe it as *belief* tout court (the lawyer may in fact believe that her client is guilty). I shall refer to pragmatic attitudes of this kind as *non-doxastic acceptances*. Non-doxastic acceptance is not subject to norms of epistemic rationality, and it may be subject to norms of procedural and agential rationality only within certain contexts. (Frankish, 2012, p.24)

However, in level 2 there are also doxastic states, which have different characteristics:

> Level 2 belief, by contrast, is a doxastic form of acceptance, which is typically formed for epistemic reasons, is context independent, and is unqualifiedly subject to the norms of epistemic, procedural, and agential rationality. (Frankish, 2012, p.24)

Delusions, according to Frankish, correspond to the level 2, non-doxastic acceptances: "Delusions are typically ascribed on the basis of the patient's avowals, rather than inference from their unreflective, nonverbal behaviour" (Frankish, 2012, p.25). However, although initially delusions are formed as level 2 beliefs, they may later generate additional level 1 beliefs. That both levels of belief exist independently from one another is how Frankish explains why some beliefs are contradictory. This characteristic is one of the key arguments used by Frankish to explain the observed compartmentalization in delusions. Also, due to their different degrees of agential

rationality, having two independent belief levels explains the poor conviction shown by some deluded subjects:

> If delusions are level 2 beliefs, then this offers a possible explanation for cases where delusions fail to influence action. The subject is strongly motivated to accept and act on a delusional claim (say, that their partner has been replaced by an imposter), but is also strongly motivated to refrain from performing certain actions which it dictates, such as leaving their partner. (Frankish, 2012, p.26)

Delusions correspond to the specific level 2 case of non-doxastic acceptances. He compares delusions to religious faith, because the latter can be adopted rather than believed in the doxastic sense. Therefore, summarizing, to Frankish, delusions are level 2 rather than level 1 because they are conscious, rather than pragmatic policies. However, they are not doxastic because they are more like adopted acceptances rather than something that motivates behaviour.

Once both levels of beliefs have been explored, it is worth doing some further analysis. It can be noted that many different characteristics of beliefs have been encompassed by this classification, and that has taken us to an unusual characterization of what constitutes a belief.

IX. General criticism of the Two-level Framework

I will first examine some of Frankish's assumptions about the status of level 1 beliefs as dispositions. It is clear from their description that these are belief-like states that are mainly focused on the subject's external behaviour.

It is not very clear what the role of an unconscious, not introspectable disposition that influences behaviour would be. It might be that they play a role like the spontaneous feeling generated when one sees a loved one, or maybe like some uncontrollable nausea when a repulsive sight appears. If this were the cse, it would not be difficult to argue that those behavioural dispositions are not a belief. But this is not what Frankish has in mind, since they seem to be more like a "working" belief, which I can act upon, but I really do not endorse. For example, one could act in a certain way only because of loyalty

to a friend, to use Frankish's own example. However, in his (2009) paper he offers what I consider a brief and not very illuminating example of what he means:

> On the other hand, we also possess a mass of beliefs which we have never thought about, which operate silently, in the background, and which we are only indirectly aware of possessing. Think, for example, of the huge corpus of knowledge about the location and function of household objects which guides your behaviour around the house. (Frankish, 2009, p. 272)

However, this explanation is not enough to rate level 1 states as unconscious.

According to Frankish's claims, level 1 beliefs are passively formed in contrast to level 2 ones, which are formed consciously and based on the evidence available. In this sense, it seems that Frankish does not mean that they are passive in the sense of "not leading to action"; what he means is that they are inert in the sense that they will not interfere with other beliefs because there is not enough evidence to sustain them. Therefore, what this claim is presupposing is that there are behavioural dispositions which are not guided by epistemic evidence. Accordingly, Frankish established that these beliefs are not subject to Bortolotti's criterion of epistemic rationality

About the claim that level 1 beliefs are holistic, this holism is only defined as a property that plays the role of providing reasons for a desired behaviour, using many of the subject's beliefs instead of a single strong belief. In this sense, Frankish's definition of holism is compatible with Bortolotti's definition of procedural rationality, and therefore it is not surprising that level 1 beliefs and procedural rationality are compatible: "If these [behavioural] dispositions include dispositions to form new beliefs, then adherence to the norms of procedural rationality will also play a constraining role" (Frankish 2012, p. 26).

In the case of agential rationality, it is the defining characteristic of level 1 beliefs, and as such, it is portrayed by them:

> In the case of level 1 beliefs, adherence to the norms of agential rationality acts as a major constraint, since the multi-track behavioural dispositions that constitute these beliefs are defined by reference to those norms. (Frankish 2012, p. 26)

Agential rationality is an important feature in most of the debates about delusions, and, although it is not stated explicitly in Frankish's account, it can be inferred that level 1 beliefs often override or screen level 2 beliefs.

In the same way as Bortolotti, Frankish is committed to a theory that might describe beliefs for both normal and deluded subjects. In Frankish's two-level framework of beliefs, there is nothing that could be remarked as being distinctively pathological in delusions; they are just clearly delimited to being a species of level 2 beliefs, and their lack of rationality does not play a defining role:

> Indeed, the two-level view actually makes it easier to argue for a version of the doxastic conception, since arguments for strong rationality constraints on belief ascription can be dismissed as applying only to level 1 belief. (Frankish, 2012, p.26)

The aetiology of delusions is clearly not the main motivation behind the two-level framework, since Frankish only sketches a very quick psychological explanation as his aetiology, saying that "perhaps patients adopt delusions because they answer some emotional or other psychological need, rather than because they are probable"(Frankish, 2012, p.27); however these claims are not taken any further.

On one side, we have some beliefs which are unresponsive to evidence, binary and only contentiously rational. However, these same beliefs are the only ones that are conscious, formed by epistemic reasons, and non-holistic but sincerely held. The combination of these features is odd, especially because these are the only mental states that in Frankish's view deserve the title of beliefs as such.

On the other side, we have these non-doxastic, behavioural attitudes, which work well with other beliefs and exhibit a high degree of agential rationality, despite the fact of being unconscious and challenging to the usual "policies". An explanation for putting together this characterization comes from Frankish's previous works, which deal with decision making theories. This might also motivate the definition of level 2 beliefs as policies which would fit with his two-level framework of beliefs, in a bit of an ad hoc way:

> Many researchers have converged on some form of dual-system theory, according to which humans have two distinct systems for reasoning and decision-making, often referred to as System 1 and System 2. (Frankish, 2009, p.273)

In a sense, Bortolotti and Frankish sustain theories with many assumptions in common. They agree on what the empirical evidence for defining the doxastic nature of delusions would consist in, since they both agree that both delusional and normal beliefs should be included in the analysis. It is worth noting that Bortolotti and Frankish both designed their theories independently, and yet they agree on the criteria that all beliefs should portray, and this can be seen in the equivalences of their rationality criteria. Also, both theories of delusions have in common that one possible conclusion from their analysis is that there is not a pathological sign that would distinguish a delusional belief from a normal one. Their main point of disagreement is that, to Bortolotti, normal and delusional beliefs show some degree of irrationality and, for Frankish, they do not. In the case of delusions Frankish is very careful to claim that they are classified as non-doxastic acceptances, based not on their rationality, but in all the other characteristics they exhibit. Therefore, delusions have the characteristic of being non-doxastic; their ontology is a flexible concept which he labels as “acceptance”.

X. General conclusion to antidoxasticism

Alternatives to doxastic theories of delusion offer a different panorama of available options that question the fundamentality of delusions as beliefs. The antidoxastic theories show a very dynamic way of approaching the contradictory nature of delusions. As we have seen through the different representative accounts, antidoxastic accounts have the following underlying assumptions:

- The content of the delusion is important- To antidoxastics, that the patient is asserting that something is going on in his mind is a factor that requires explanation with priority over other empirical characteristics of the

phenomenon. Other philosophical accounts, for example, phenomenological stances, dispute that the content of the delusion is important[5].

- The term belief is not sufficiently well defined as it is- Antidoxastics tend to propose new alternatives that might take the role of beliefs. Therefore, they assume that the concept is not capable of accommodating the phenomenon because it is ill defined to do this task.
- Delusions are closer to different propositional attitudes distinct from beliefs- Because delusions are assumed to have content, a propositional attitude is a viable option to categorize delusion. However, to antidoxastics, the concept of belief does not describe adequately the phenomenon in some crucial aspects.
- Explaining the empirical data of the phenomenon takes priority over the categorization of the phenomenon in a familiar concept- Between having the choice of making the concept of belief more flexible, and explaining the puzzling bits of the phenomenon with a new framework, antidoxastics choose the latter.
- That the delusional proposition has meaning is assumed- The question of meaning in the utterance is not relevant to the ontology of the phenomenon. Meaning is assumed unproblematically, since the meaning can be deduced from the new framework. For example, if delusions were an imagining, the meaning of the utterance is such that one is only entertaining the idea of such and such.
- It implies that it is not possible to empathize with such alien, pathological states which are unknown to us- Since antidoxasticism is proposing a new, pathological, ad hoc, state it is the case that empathic understanding cannot be achieved fundamentally because there is nothing similar in our everyday life.
- Aetiology is important but not central to the ontology- The causes of the phenomenon sometimes help to decide which ontology has the potential of explaining some empirical data, but the question of the true nature of delusion is independent of the former.

However, these alternatives have been unable to refute the main claims of doxasticism, or to give a solution to all the problems encountered in descriptions of delusion. Anti-doxasticism responds to the worry that delusions cannot be beliefs by suggesting that they are some other non-doxastic propositional attitude. Delusions are an instance of

[5] This characteristic of phenomenological accounts is explained in the next chapter.

some other propositional attitude which antidoxasticism will spell out to shed light on the nature of delusions.

Their strategy, however, faces the following dilemma: either delusions are compared to a familiar propositional attitude, or they compare delusions to non-standard propositional attitudes. Then:

1) A familiar propositional attitude promises to shed light on delusions, because we already know the role of the propositional attitude in question. But the problem with this approach is that delusions seem unlike familiar propositional attitudes. In effect the same charge antidoxasticism raises against doxasticism can be directed back against it. For example, Currie cannot say that delusions are just imaginings. He has to say that they are imaginings mistaken for beliefs, a second-order cognitive hallucination (believing that one believes p). Additionally he then suggests that this causes a first-order belief. So he moves off the first horn and onto the second. Summarizing, in this horn, although antidoxasticism makes use of familiar attitudes, these cannot become unfamiliar enough to be explanatory.

2) In non-standard propositional attitudes the objection is that the ad hoc nature of the propositional attitude means no light is shed on delusion because we do not understand what it would be to have such a propositional attitude. For example, a bimagining; a delusional stance; a cognitive hallucination etc. This means that we end up with something like the sui generis type.

The question now is whether another type of explanation would be adequate to overcome the dilemma above. Therefore I will continue with the contrasting option: doxasticism about delusions.

CHAPTER III- DOXASTIC ACCOUNTS OF DELUSIONS

I. Introduction

In the literature, reports that described delusions as mistaken *beliefs* first appeared around the 19th century, along with modern definitions of belief (Berrios, 1991). At the same time, their rival views appeared: antidoxastic accounts. These proposed alternative theories to the belief status of delusions. However, in the early 20th century, when Jaspers' definition of the features of delusion prevailed, the notion of them being fixed beliefs became paradigmatic. It was within the recent debate about the nature of delusion in the 1990's that antidoxastic views emerged again.

Nowadays, the arrival of the antidoxastic movement changed the landscape of the philosophy of delusions in the sense that the opposing doxastic views did not return to the previous full-belief conception; the modern approach, as we are about to see in this chapter, advocates instead a more modest version of belief. However, defenders of the belief approach have very consistent views that are compatible with our modern intuition of what a delusion is.

In this chapter I analyse three different doxastic positions: Bayne and Pacherie's stance of delusions, in which the latter can fit in a more encompassing definition of belief; Bortolotti with the proposal of a continuum of irrational beliefs; and finally, Reimer's view of delusions as beliefs which are not paradigmatic. These are representative examples of philosophical moves that still aim to capture the nature of delusion within the concept of belief.

II. Bayne and Pacherie's defence of doxasticism in delusions

Tim Bayne and Elisabeth Pacherie (2005) argue for an account of delusions that defines the term "belief" in such a way that most delusions would fit into it. In order to do so they set apart the worries generated by one of the most influential antidoxastic accounts, discussed in the previous chapter: Currie's assertion that delusions are "cognitive hallucinations" reflected in an imaginative state that the subject mis-identifies as a belief. Since Currie makes explicit that his account works via a deficit in

the monitoring of mental states, Bayne and Pacherie call Currie's account the "metacognitive model".

a) Against Currie's "metacognitive" model of delusions

In their work 'In defence of the doxastic conception of delusions' (2005), Bayne and Pacherie summarize the core premises of Currie's metacognitive model, as follows:

- Delusional patients who seem to believe P do not actually believe P.
- Delusional patients who seem to believe P actually imagine P.
- Delusional patients who seem to believe P believe that they believe P.

And they acknowledge the following virtues of the model:

1) It accounts for the objection that delusions are incoherent (the self-defeating objection) because in imagination nothing can be truly incoherent regarding its content.
2) It can account for delusions not being based on evidence, because imaginings do not have to be either.
3) It accounts for having inconsistent beliefs, because imaginings do not have to cohere with the person's other beliefs.
4) It accounts for delusions failing to engage with the patient's affective life.
5) It also accounts for the patient's assertion that they believe in a specific state of the world.

However, Bayne and Pacherie offer a number of objections against Currie's account. The first is that Currie's explanation of what differentiates a belief from an imagination consists in three unrelated criteria (Bayne and Pacherie, 2005, p.168):

(i) the relations that they have to perceptual states (input).

(ii) the consistency constraints that they are subject to.

(iii) the relations they have to action (output).

Bayne and Pacherie look at these criteria and object to each one separately, as I summarize as follows:

- **Against (i) the relations to perceptual states**

An objection against Currie and Ravenscroft's (2002) paper concerns the generation of beliefs and imaginings. The distinction, in that paper, is drawn like this:

Sometimes they say that imaginings are actions—they are doings—whereas beliefs are not actions, they are things that happen to us (Currie and Ravenscroft, 2002, p.25). In other places, they make the weaker claim that imaginings are autonomously generated, whereas beliefs are formed 'in response to perceptual information, or by inference from other beliefs we already have' (Bayne and Pacherie, 2005, p.169).

But this distinction is not sufficiently clear, Bayne and Pacherie argue, since on one hand imagination is not necessarily under the control of the will, and on the other, forming and maintaining a belief can be an action. Even the distinction between the endogenous and exogenous[6] sources of imaginings and beliefs is problematic because, under the influence of emotions, beliefs can be endogenously generated too, such as in wishful thinking and self-deception (Mele, 1993). Also, it is plausible that some beliefs might be innate as well. This being at least possibly the case, Bayne and Pacherie state that top-down accounts[7] are coherent, although not plausible:

> If these non-evidential routes to belief are possible, there seems little reason to rule out the possibility of belief being formed as the direct result of brain damage. Beliefs are typically formed in response to perceptual information or by inference from other beliefs, but we see no reason to think that these are the only routes to belief. (Bayne and Pacherie, 2005, p.169)

Therefore, the relation that Currie proposes between the input (that is, the perceptual state) and the delusion cannot be as Currie defends in the first place because the formation of the delusion is unlike the top-down formation of imaginings.

---

[6] Definitions extracted from (Di Tella et al., 2004): Endogenous Beliefs: Research in sociology and psychology suggests that individuals sometimes engage in belief manipulation to improve their expected utility when the costs, for example in terms of the mistakes that this induces, are not too high and when there are enough benefits, for example in terms of the effort they induce.[…]
Exogenous Beliefs: In this view beliefs are formed as a by-product or consequence of an agent's activities and interaction with others in society. It includes the case of people who would actively desire to know the value of some parameter in society (e.g., how much can you trust others or how important is money for happiness), and come round to their beliefs through their own experience and the observation of other people's experience.

[7] A top-down account of delusion implies that the mind of the patient forms the flawed propositional attitude first (for example, a belief), and then, such propositional attitude influences the patient's perception or experience. Accounts that endorse top-down mechanisms usually propose some sort of cognitive bias, or a flawed background of beliefs (Cfr. Hohwy, 2004)

- **Against (ii) the consistency constraints that they are subject to**

First, Bayne and Pacherie state that one important remark has to do with Currie's "consistency constraint", which is the following:

> Imaginings seem just the right things to play the role of delusional thoughts; it is of their nature to coexist with the beliefs they contradict, to leave their possessors unwilling to resolve the inconsistency, and to be immune to conventional appeals to reason and evidence. (Currie and Jureidini, 2001, p. 160)

Bayne and Pacherie first draw a distinction between a *normative* and a *constitutive* conception of the consistency constraint. The normative conception states that one ought to have beliefs which are consistent among themselves, and an agent should take this in account when revising his beliefs. The constitutive conception is that one cannot believe in P if there are other things also believed inconsistent with P. In the case of delusions, Bayne and Pacherie argue that it is the latter option, the constitutive conception, that can shed some light on the distinction between imaginings and beliefs, since the normative conception only says that beliefs and imaginings ought to be different. However, this option can be ruled out too as overly restrictive, due to the presence of counterexamples such as self-deception and compartmentalization.

Second, Bayne and Pacherie defend doxasticism against the apparent counter-example of epistemic inconsistency. Currie states that something particularly damaging to the doxastic account is that delusional subjects are indifferent to inconsistencies in their beliefs even when asked about it. That means that their awareness of the inconsistency, and their unwillingness to correct their thoughts after that, is what undermines the status of delusions as beliefs. Currie's argument continues by observing that, due to the fact that patients acknowledge the tensions between their own beliefs, delusions cannot be treated in a doxastic manner. Bayne and Pacherie argue that patients often revise their delusional beliefs, and try to give the best arguments available in order to defend them. Even more, patients are not indifferent to the aforementioned tension, and that is an even stronger argument towards saying that delusions are beliefs.

- **Against (iii) the relations they have to action**

Currie claims that imagination tends to have fewer effects on behaviour than full-fledged beliefs do. The first problem with that is: why does the patient's verbal behaviour seem to suggest the contrary? Currie's answer (drawing on Graham and Stephens, (2007)) is that the patient only believes that he has the belief. That is, he has a second-order belief instead of actually believing (at the first-order) the content of his utterance. Against Currie's assertion (and against Graham and Stephens', consequently) Bayne and Pacherie put forward the following two objections:

In principle, the distinction between first-order beliefs and second-order beliefs is about their functional role: first-order beliefs are not about other of the subject's mental states, whereas the second-order states are precisely *about* first-order mental states. Therefore, verbal assertion of delusions implies that they are explicit beliefs, and there is no evidence to state that it is a second-order belief rather than a first-order one.

Secondly, Currie would like to grant both (first-order and second-order beliefs) the consistency and the action-guidance constraint. If the patient has the <belief that P>, this should have an impact on his behaviour (both verbal and non-verbal) other than having straightforwardly the belief that <P>. Therefore, if the encapsulation of <P> is a problem for the doxastic accounts, it is also a problem for the meta-doxastic account of having <the belief that P>. In other words, Currie cannot explain why first-order states do not guide action adequately on some occasions, and neither can he explain why second-order states cannot do so, in any case.

b) Bayne and Pacherie's positive doxastic account

As I described earlier, Bayne and Pacherie's (2005) account argues for a definition of belief that will encompass most delusional states rather than trying to fit all delusions on a doxastic account. They acknowledge that they based their model of belief on Cherniak (1986) and Schwitzgebel (2002). Also, they concede that the definition of belief is far from homogeneous; even more, definitions of belief that are based on rationality and rationality constraints, like (Davidson 1973, Quine, 1960, Dennett, 1971), actually "obscure the multi-dimensional nature of belief" (Bayne and Pacherie, 2005, p.179). To them, there are other important constraints on belief ascription.

The first of the belief ascriptions that they argue for, based on Cherniak's criteria, is that rationality is constrained by memory, mental processing time, and energy; therefore, it is not close to the ideal rationality (such as Davidson's) maintain. Therefore, only minimal rationality constraints apply most of the time. This means that an agent should only undertake *some* of the appropriate actions, make *some* of the appropriate inferences, and eliminate *some* of the inconsistencies that may arise among beliefs as dictated by the norms of rationality in order to have a doxastic state.

The second belief ascription, from Schwitzgebel, claims that beliefs are context-dependent in several ways. The portrayed dispositions are influenced by long-term memory, external context, and the affective life of the individual. Beliefs are ascribed in clusters activated in certain situations, called stereotype beliefs. Therefore, according to Bayne and Pacherie, social psychology has shown that this and Cherniak's criteria for belief ascription are more appropriate to describe beliefs than the rationality criteria.

c) Counter-arguments in favour of the doxastic account

Bayne and Pacherie take a stance against three main antidoxastic arguments. The first one is about the doxastic content, the second one is about its evidence, and finally, about commitment in beliefs.

One of the general claims in doxasticism is that the content of some delusions is not logically possible. Bayne and Pacherie argue that using this point to defend antidoxasticism constrains the notions of belief and logical possibility. To defeat that, one has to observe that it is conceivable to believe something that is a logical falsehood. For example, it is possible to believe that one person has a nuclear power plant inside his body even though this is necessarily false because a nuclear plant cannot fit inside a body. They also argue that, even in the case of apparently blatantly false delusions, there may be other interpretations available because, in other contexts, or "language games", that claim might have some sort of meaning to the believing person. For example, in the Cotard delusion, when the patient utters that she is dead, she means "dead "not necessarily in all contexts; it could be in the religious, folk-psychological, subjective realm that the person is in, and not strictly in the biological sense:

> In his use of words, a delusional patient may break the rules governing one of these intersecting language games while retaining the meaning his words have in other language games. (Bayne and Pacherie, 2005, p.182)

One of the marks of delusion is that it is held in the absence of evidence and in the face of counter-evidence. Bayne and Pacherie take here a strong and straightforward stance; "We deny that there is a constitutive connection between belief and evidence" (Bayne and Pacherie, 2005, p.183). Their argument is that both scientific and folk psychology propose that beliefs can be innate and formed by motivation mechanisms. Therefore, it seems plausible to have a delusional belief by having received some sort of brain damage. And these beliefs would still guide theoretical and practical reasoning as other beliefs do. To argue for the applicability of doxastic-like evidence in the case of delusions they suggest that, although delusional patients are not able to produce evidence that is available to others, they often have first-person evidence for their beliefs.

Another key antidoxastic argument is that delusional patients fail to demonstrate the kind of commitment to the content of a delusion that one would expect from someone who believed it. Patients fail to manifest the expected behaviour, showing a lack of the proper affect, theoretical, and practical reasoning that is expected or typical of someone with that belief.

To Bayne and Pacherie the point is, rather, that patients do not show the *stereotypical* responses to having that belief, and there is a number of reasons of why they might not do so. The first is "if, as the empiricist claims, monothematic delusions such as the Capgras delusion are grounded in unusual perceptual and affective experiences, then the patient's belief might be continually reinforced" (Bayne and Pacherie, 2005, p.184). Thus the subject's conduct and the other effects of the belief are also affected, resulting in a merely apparent lack of commitment to the content.

Bayne and Pacherie also observe that there are certain delusions whose content cannot be verified. They give the example of thought insertion, where it is not obvious how to prove or disprove that external thoughts are being inserted into the patient's mind. In any case, thought insertion is not accepted as a natural phenomenon, but it is not logically impossible. Also, in delusions where emotion is relevant, Bayne and Pacherie suggest that it is not clear what the appropriate response would be; clinicians usually expect a certain stereotypical belief, but that is not the only rational, available option.

In other pathological cases, like depression, the subject disassociates completely from her emotions, but her beliefs do not lose their doxastic status. In cases with Cotard patients, depression has been linked to the condition, and therefore it might be the case that "some delusional patients have the subjective experiences associated with certain emotions even when they lack their behavioural manifestations" (Bayne and Pacherie, 2005, p.184). This implies that we cannot know if certain individuals are bound to have certain emotions given their subjective circumstances.

Finally, Bayne and Pacherie argue about delusion-generated action as commitment. To start with, there are clear, not so rare instances where the delusional patient acts on his beliefs. Secondly, one has to observe that, in any case, action is not only motivated by cognitive states, since it has to do with affect too. They also claim that doxastic states are context-sensitive, as is the case with Capgras patients, which manifest their abnormal behaviour only when seeing the relevant "impostor" in front of them. They suggest that this could be due to the type of brain damage suffered:

> We might liken this behaviour to the field-dependent behaviour that one sees in frontal lobe patients, who are unable to inhibit their pre-potent responses to environmental cues. Perhaps some delusional patients are doxastically field-dependent. (Bayne and Pacherie, 2005, p. 185)

Bayne and Pacherie's final remark concerns the implications for treatment in classifying delusions as doxastic or antidoxastic. In the case of cognitive behavioural therapy (CBT) it is clear that treating the delusion as a belief is necessary in order to try to convince the patient that he has a belief which should be attenuated as much as possible. But in the case of Currie's metacognitive account, it is far from clear what sort of therapy would help, since the patient does not have the belief that <P>. The aim would be to try to get rid of the imaginative anomalous state, given the fact that it is affecting the patient's quality of life, even if it is classified as an imagining; and it is not clear how this could be achieved.

d) Objections to the (dispositionalist) doxastic account of Bayne and Pacherie

Maura Tumulty, in her (2011) paper criticizes Bayne and Pacherie's position, calling it 'dispositionalism'[8], because it tends to play down the notion of belief to accommodate it to possible delusional instances. In other words, in dispositionalism, context matters to understanding.

However, Tumulty challenges that dispositionalism is a valid model: "The trick here [in Bayne and Pacherie's account] is to show that the account appealed to in order to count delusions as beliefs is really still an account of *belief*"(Tumulty, 2011, p.598). She says that, based on Schwitzgebel's concept of dispositionalism, the context dependence is useful to point out how a belief should be communicated by the speaker in order to be understood according to what the subject actually thinks and to avoid misunderstandings. Tumulty's main objection to Bayne and Pacherie is that they fail precisely to give an account of belief because their position is too ad hoc*:* "Dispositionalism can't distinguish beliefs from other propositional attitudes without making assumptions about believers that aren't true of deluded subjects" (Tumulty, 2011, p.598).

Bayne and Pacherie claim that patients might fail to act on their belief due to the fear of being labelled irrational by other people. Tumulty says that such an attitude *excuses* the patient from his behaviour, but it does not *explain* why she has that disposition as a *belief*. Tumulty's objection is not necessarily that delusional patients do not just choose to not act on their delusion, but that they do not act because the belief does not have the impetus to make the subject perform the action. In other words, it cannot be known if subjects really *inhibit* their dispositions to act according to their delusions or if they just *lack* them.

The problem of dispositionalism in monothematic delusions is the following: one has to find a cognitive deficit that it is severe enough to form the delusion, and that is restricted only to the delusion without affecting the surrounding attitudes towards it (if you want to still propose it as a belief). That is, the problem in monothematic delusions is that the cognitive deficit seems to be selective: for example, a patient with Capgras will only be

[8] Dispositionalism, according to Schwitzgebel, can be defined as: "[A] dispositional account (…) treats believing as nothing more or less than being disposed to do and experience certain kind of things" (Schwitzgebel, 2002, p. 250).

hostile towards his wife whilst the rest of his cognitive capacities remain unaffected. Therefore, Tumulty observes that, granted that the abnormal perceptual experience will prevent the subject getting to the non-false belief, how is it that the subject still does not consider more rational alternatives to his delusion, such as achieving the best possible belief? Bayne and Pacherie would respond that the subject's belief is personal and subjective, and that they will try to test their hypothesis, but their abnormal experience will tell them the same delusional answer every time. Tumulty's response is that this can be granted in the case of Capgras since the expert testimony about the fact that his spouse has been changed is the deluded subject himself, because according to the subject, he is the one who knows his partner better (or at least, better than external witnesses, who would not notice the slight differences the impostor has). However, Tumulty states that this can only be applied to this specific case, and the point would not apply to other delusions which might have a wider pool of experts.

Bayne and Pacherie also argue that some delusions are unfalsifiable, such as thought insertion. Tumulty expands what they might have wanted to say:

> [D]eluded subjects are in fact carrying out plausibility comparisons for their delusions. They are, that is, manifesting some of the cognitive dispositions relevant to belief in those delusional contents. But we don't realize that they are, because the notions of plausibility and causation to which they now appeal are quite alien to our own, grounded as they are in very different experiences. (Tumulty, 2011, p.606)

Tumulty argues that, if deluded subjects are reasoning so badly as to arrive at such conclusions, given the fact that they are trying to reason, the subjects' notion of causation would be practically non-existent.

In the case of subjects lacking the proper behavioural attitude because of the relevant lack of affect in the belief that Bayne and Pacherie suggest, Tumulty observes that they are proposing a disruption in motivation to explain the aboulic[9] behaviour. She argues that dispositionalism only accepts this excuse in the form of "lack of consciousness", and that is something that Bayne and Pacherie would like to avoid. Also, folk-psychology

[9] Aboulic means that the subject shows inability to act or make decisions.

accepts that one has the dispositional belief only if one has the disposition to act in the relevant ways: “On the dispositional model of belief, we might say, beliefs just can’t be motivationally inert” (Tumulty, 2011, p. 608). In any case, Tumulty says that lack of motivation looks just the same as the disposition of not having the belief at all. Their response might consist in asserting that emotional and affective responses are not “constitutive elements of the belief stereotype” (Bayne and Pacherie, 2005, p.184). In that case, other dispositions would have to become stronger, such as rationality and practical reasoning, but deluded subjects do not seem to be better at these tasks either. One might be tempted to attribute the doxastic state to those beliefs which are acted upon. In one study, 96% of patients with persecutory delusions have acted according to that belief (Freeman et al, 2007). However, most of the time deluded patients fail to show those dispositions, or do so inconsistently. Regarding misidentification syndromes, 50% of the patients fail to act according to their delusion (Forstl *et al.,* 1991). So, there is still a large percentage of delusional people who do not act on their beliefs. So, in those cases, Bayne and Pacherie still have to explain why should they have that belief. Tumulty argues that dispositionalism would not work because it does not fit the judgement requirements: “It doesn’t express all-things-considered judgments that a deluded subject really believes his delusional content, despite his having some dispositions that fit the stereotype for imagining them instead” (Tumulty, 2011, p.614). Tumulty’s final remark is that Bayne and Pacherie’s stance has the fundamental characteristic of linking belief-ascription to clusters of dispositions; it avoids making any particular disposition *necessary* to belief, and therefore allowing anyone and anything to have a belief.

All in all, Tumulty’s objection amounts to Bayne and Pacherie’s inability to explain the pathological nature of delusion by means of their doxasticism: it plays down the cases where delusions show their abnormality and cannot explain the overall mechanism in which delusions work. Doxasticism works well for the cases in which the functional role of a delusion is similar to a belief, but there is no explanation of why delusions sometimes are like beliefs and sometimes not. In other words, doxasticism seems to be forced upon delusions; it is too ad hoc. In the end, the doxastic debate disputes the nature of delusions by arguing about the essential and non-essential features of belief. In any case, the problem is that the features of belief are not well defined enough to be conclusive.

III. Bortolotti's modest doxasticism about delusions

In her (2010) book, Lisa Bortolotti aims to assess the rational status of delusions. Her main claim is that delusions are in a continuum with non-delusional thoughts *as beliefs*. Although delusions are on the irrational side of the spectrum, they are nevertheless, ontologically identical to normal beliefs, except for their *degree* of rationality. Bortolotti also wants to establish that normal beliefs are not as rational as one would initially think either. So, Bortolotti's strategy has two elements. One aim is to prove that irrationality does not entail not being a belief. The other element consists in trying to downplay our notion of rationality in normal beliefs. Her target is clear: to make the irrationality gap between delusions and normal beliefs as small as possible in order to make way for her doxastic account. In her defence of doxasticism, Bortolotti takes into account three different types of rationality: Procedural, Epistemic and Agential. I will discuss her arguments according to each account in the following sections.

a) Procedural rationality

Procedural rationality comprises the claim that, in order to be rational, beliefs should be well integrated with the subject's other beliefs. Delusions seem to violate this norm. To accord with this procedural rationality constraint, subjects should not make obvious mistakes in deductive reasoning and they should obey basic inferential rules among beliefs and other inferential states. Philosophers such as Dennett (1971) and Davidson (1982) claim that having this type of rationality is necessary to ascribe beliefs to a subject. However, some counterexamples have been put forward against such a constraint, such as Stich (1984), where he argues that there is nothing incoherent about the idea of people having, for instance, inconsistent beliefs, because one can make sense of this. Hence rationality cannot be a pre-requisite of intentionality. Stich quotes empirical psychological literature on ordinary irrationality (Stich, 1984, p.52):

> In a similar study with alarming implications for public policy judgments, Slovic, Fischoff, and Lichtenstein (1977) showed that subjects estimate the probability of a compound sequence of events to be greater than the least likely of the events in the sequence. It is disquieting to speculate on how large an impact this

> inferential failing may have on people's assessments of the chance of such catastrophes as nuclear reactor failures which require a number of distinct events to occur in sequence. (Slovic and Fischoff, 1978)

Therefore, in order to counter this sort of irrationality, a strategy called the "background argument" has been devised by Davidsonian philosophers. It states that:

> Only within a belief system that is largely (procedurally) rational can belief status be granted to belief-like states that fail to meet the standards of (procedural) rationality. Dissonant beliefs and beliefs that result from mistakes in deductive reasoning or bad inference may be characterised as beliefs, but only against a general background of (procedural) rationality. (Bortolotti, 2010, p.63)

Bortolotti grants that delusions can be quite puzzling with respect to this type of rationality because patients will express contradictory beliefs, for example the Cotard delusion in which one states that one is dead. However, delusional subjects also usually try their best to make many of their beliefs cohere with their delusions, and they can be found offering reasons for their bizarre thoughts. Also, subjects sometimes behave as if they do believe the content of their delusion. More generally, Bortolotti argues that the procedural rationality constraint is too much for any belief, even in normal people. Psychological literature about human reasoning shows that, in the normal population, the vast majority of subjects have failings in procedural rationality by making systematic mistakes in probabilistic, statistical and deductive reasoning (Stein, 1996; Samuels et al., 2002). Therefore, Bortolotti claims:

> Although there are significant individual differences in reasoning performance, and certain reasoning biases may be more pronounced in deluded and high risk samples, there is a concrete possibility *that delusions and beliefs are continuous in their procedural irrationality* and that, with respect to beliefs, the conditions for intentionality and those for rationality should be kept distinct. (Bortolotti, 2010, pp.77-78)(my italics)

The "background argument" requires the subject to not have inconsistent beliefs at the same time, but, Bortolotti argues, sometimes normal subjects seem to fall into this specific type of error too.

The first of Bortolotti's examples is *preference reversals*, in which one can show that subjects usually do not grasp the logical inferences coming from their own choices. One of the typical examples of this is when, given a certain choice between A or B, the subject prefers A; however, if A and B are described in a semantically different way, which has no effect on the logical value of A or B, the subject sometimes prefers B. One example[10] of this behaviour is the Tversky and Kahneman (1981) experiment, in which two groups of normal subjects were presented with the same problem but with a different phrasing, one more positive than the other. Although the choices given to both groups were equally good, when the same alternative was phrased in a negative way to one of the groups, that choice lost a significant percentage of subjects compared to the other group.

These logically incompatible preferences are mostly present when there are ethical, or life and death, issues involved (Bleichrodt and Pinto Prades, 2009). This phenomenon concerns preferences rather than beliefs, but Bortolotti states that it is still a clear failure of procedural rationality.

The second of Bortolotti's examples is that normal subjects usually do not make the correct inferences in simple probability tasks. One example is another Tversky and Kahneman (1983) study, where subjects violate the logical conjunction rule. However, Hertwig and Gigerenzer (1999) showed that the subjects' performance improves if the problem is not put like a plausibility problem and more like a probability one. Bortolotti concludes that normal people are prone to formulate and maintain beliefs that are not necessarily well-integrated with one another. Another case in which incompatible beliefs arise is the phenomenon of dissonant beliefs, which is better exemplified by superstitious beliefs: a person can believe that his green necklace brings him luck for the exams, but maintain at the same time that he has to study for the exam in order to pass. One important difference between deluded subjects and normal subjects regarding the bad integration of beliefs is that, apparently, the latter acknowledge their reasoning mistake when it is pointed out to them, but delusional people still maintain their

[10] Other examples are Tversky and Thaler (1990) and Stalmeier et al (1997).

delusional belief in the same case. Bortolotti calls this ability "recovery after Socratic tutoring" (Bortolotti, 2010, p.86). In other words, normal subjects are disposed to restore the good integration of beliefs whilst deluded subjects cannot. Bortolotti argues, however, that normal subjects are highly resistant to giving up their own incompatible ideas such as self-serving beliefs, superstitions,[11] and the fact they do not use logical conjunction properly.

Davidson explains these type of beliefs within a phenomenon called compartmentalization (Davidson, 1982). In Davidson's view, the limited resources in attention and memory force the subject to store her beliefs in different mental compartments which prevent the subject from having two incompatible beliefs at the same time. When the subject does not have compartmentalization, she can recognise incompatibility between beliefs and restore procedural rationality. However, Bortolotti argues that compartmentalization can be seen as an option to understand delusions as beliefs as well. The case of monothematic delusions, where the belief is highly circumscribed and the subject's other beliefs are unaffected, is a clear example of extreme compartmentalization. On the other hand, in polythematic delusions, beliefs tend to be better integrated and less dissonant with the subject's other beliefs, as in delusional paranoia. "This suggests that the correlation between compartmentalization and the presence of inconsistencies *supports* rather than *undermines* the continuity between delusions and beliefs" (Bortolotti, 2010, p.89). But, if compartmentalization is a valid explanation is a different question altogether. Bortolotti argues that, in principle, there is no need to appeal to compartmentalization because normal subjects fail to restore the rationality of their beliefs when inconsistencies are shown. Also, subjects with delusions seem to take into account counter arguments as well, and they try to make their beliefs coherent as much as they possibly can; therefore, the compartmentalization theory is not really necessary.

There is an additional worry about procedural rationality in the case of delusions which concerns the relationship between rationality and intentionality: even though it is possible to ascribe beliefs to people who have irrational behaviour, the belief one ascribes to them might not be determinate in content; therefore, there is a necessary relation between intentionality and rationality. Bortolotti affirms that this does not have

[11] Tversky and Kahneman (1983)

to be the case, because there are other ways to explain the difficulties that the interpreter may have when ascribing beliefs with a determinate content; for example, the interpreter might ignore some of the subject's background beliefs, or maybe some causal relations in the environment. Therefore, Bortolotti concludes, rationality does not have to play a major role in belief ascription, although it is a useful heuristic when interpreting fully competent human adults (Bortolotti, 2010, p. 111). In these cases, the subject's behaviour might be intelligible even when it violates the epistemic, agential and procedural norms of rationality, but still, can be ascribed with the belief.

b) Epistemic rationality

The most influential arguments against doxasticism usually attack the epistemic features of delusion. In order to respond to those objections, Bortolotti analyses two forms of epistemic rationality in delusions and in normal beliefs: good evidential support and responsiveness to evidence. She does not deny that delusions have faulty epistemic rationality, but she tries to argue that this fault is not different in kind to other beliefs:

> Endorsing a hypothesis with low probability may be a sign of epistemic irrationality, but does not necessarily constitute a knock-down argument against the view that the content of the delusion is genuinely believed. (Bortolotti, 2010, p.121)

Delusions, she claims, might also be just performance errors. According to Davies and Davies, a subject with delusions conforms to the following:

A. She accepts a hypothesis which does not have high probability given what she already knows.
B. She underestimates the probability that the evidence would occur even if the hypothesis were false.
C. She ignores alternative explanations for the evidence (Davies and Davies, 2009).

Bortolotti analyses the latter and states that A) is a violation of procedural rationality as explained in the section above, but B) and C) constitute serious violations to epistemic rationality, and these cannot be overcome.

So, what is the factor that makes deluded subjects worse in epistemic reasoning than normal subjects? Findings are controversial. It has been proposed that delusional patients perform just as the normal population regarding reasoning biases and deductive reasoning. However, one study that has found a significant difference is Huq *et al* (1988), where subjects have been found to jump to conclusions in Bayesian tasks, and in fact, they seem to be able to overcome a conservative bias in the normal population, therefore, being better than normal, control subjects in those specific probability tests. However, Fine et al (2007) argue that this only affects the acceptance of the delusion, and not the formation on the basis of the available evidence. A different study also shows that delusional people, especially those with a severe psychopathology and those with paranoid delusions, tend to have a stronger attributional bias (Bentall and Kaney, 1996; Kinderman and Bentall, 1997)- an attributional bias "is a cognitive bias that refers to the systematic errors made when people evaluate or try to find reasons for their own and others' behaviours"-(Heider, 1958, p.322). Still, that is not conclusive evidence that delusions are not beliefs, because some studies have shown that, when confronted with counterevidence, some subjects can lose confidence in their delusion and can sometimes abandon it through therapy (Bret-Jones et al, 1987; Chadwick and Lowe, 1990).

In contrast, Bortolotti claims that the normal population also exhibits biases that affect their epistemic rationality. People sometimes attribute causation to events that preceded another one in time, although there does not have to be necessarily a real causation in those cases (Lagnado and Sloman, 2006). Another important epistemic bias is that people have a tendency to evaluate data on the basis of their preferred theory (Chinn and Brewer, 2001), showing an additional confirmation bias. Also, there is a tendency to attribute prior thoughts to actions, even if the action was unconscious or involuntary (Wegner and Wheatley, 1999). It is also the case that people tend to think about neutral actions as willed or intentional (Rosset, 2008). Bortolotti thinks that these sorts of biases play an important role in the formation of delusions:

> Can we identify elements of continuity between the epistemic irrationality of delusional reports and that of ordinary beliefs? Although differences remain, the weakness of causal reasoning that lead subjects to accept theories with little evidential support, and the inaccurate attributions of blame and responsibility

> due to the intentionality bias are observed in subjects with schizophrenia too. The subject's interpretation of the desires and intentions of others are the origin of persecutory delusions, and the formation or justification of the delusional state often depends on attributions of meaning to physical and mental events relevant to the subject. (Bortolotti, 2010, p.145)

She also recalls that normal people tend to have a high influence from self-serving biases, such as signalling -signalling is a kind of self-deception in which subjects deceive themselves to supposedly achieve more desirable results, ignoring the available evidence- (Quattrone and Tversky, 1984), and talking about their past experiences in a more favourable way (Miller and Ross, 1975). Another clear epistemic violation of rationality, very common and hard to eradicate, consists in unrevisable beliefs such as racial prejudice, religious beliefs, conservatism and self-fulfilling prophecies.

What Bortolotti is trying to show by pointing out all these biases and faulty reasoning is that, although delusions are not excusable as a temporary mistake in the light of the background argument, many normal beliefs are also badly supported by the evidence and therefore fail to satisfy the epistemic rationality constraint as delusions do. So, in the end, she concludes that delusions and normal beliefs do not satisfy the epistemic rationality criterion.

Her next question is whether the agential rationality criterion can be what differentiates delusions from normal beliefs.

c) Agential rationality

Bortolotti (2010) affirms that there are two different types of agential rationality when considering the case of both delusions and normal beliefs: One concerns the relationship between a belief and action guidance, and the other, the relationship between a belief and reason giving.

1) Beliefs and delusions through action guidance

Delusions have been associated with pragmatic failure rather than a cognitive one, such as in Fulford (1998). Bortolotti suggests that there is a straightforward way in which

delusions affect the subject's behaviour: they necessarily have to affect negatively the patient's conduct in order to be diagnosed as delusions in the first place. This negative behaviour can usually be found in the patients' lifestyle in terms of stress, depression, preoccupation, social withdrawal, and so on. However, agential rationality demands more from the patient. It demands that the subjects' behaviour allows it to be described as intentional. Therefore, according to Bortolotti, in order to have agential rationality in delusions, the subject must act as someone who has that specific delusional belief and acts according to it.

To say that delusions guide or affect behaviour is, to say the least, controversial. On the one hand we have cases in which the delusional patients have become violent and have harmed or killed the object of their delusion (Silva et al, 1994; Ames, 1984; Wessely et al, 1993). Also, some psychologists emphasize the extensive behavioural effects that delusions seem to have on the subjects' conduct (Kopelman, 1987). On the other hand, we have reports from patients that show a conduct of 'double book-keeping' where they say they believe in something but do not act in accordance with that belief (Sass, 2001; Bleuler, 1950). A further difficulty in ascribing beliefs through behavioural states is that sometimes the belief is too abstract or too complicated to be described just by actions; for example, if a person has the belief that someone inserts thoughts in her head, how is she supposed to act?

Keith Frankish (2009) argues that because of their tendency to be behaviourally inert, delusions should be seen as mere *acceptances*. Bortolotti argues, however, that the inertness of delusions is a "philosophical myth rather than a fact needing explanation" (Bortolotti, 2010, p.171):

> As we argued, delusions might fail to guide action in some of the appropriate circumstances, but do always manifest in behaviour, or they could not be detected, diagnosed, and considered as harmful to the person reporting them. Moreover, when action guidance is compromised, we need to remind ourselves that often delusions are accompanied by other pathologies that are likely to affect the subject's motivation to act. (Bortolotti, 2010, p.171)

She also points out the agential irrationality conforming to action guidance in normal beliefs. Her first argument considers hypocrisy, where hypocrisy in normal beliefs

concerns those situations in which normal people state that they have a certain attitude towards a particular issue, but in the end they act in a way opposed to that attitude due to special circumstances. This seems to be the case in studies that involve, for example, the use of condoms, where the population usually endorses their use, but when asked if they make use of them, people said that they actually did not, because they think of them being a "nuisance" (Aronson, 1999, p. 114). Her second point considers poor self-prediction, where normal people assert a certain attitude, but in the end, due to "special circumstances", they consistently do the opposite (Ajzen, 2005). This leads Bortolotti to conclude the following:

> [S]ome beliefs have the capacity for guiding action, but action guidance should not be listed among the desiderata for rational beliefs, and not among the necessary conditions for belief ascription. (Bortolotti, 2010, p.172)

2) Beliefs and delusions through failure of reason giving

Central to the notion of agential rationality is the capacity to give good reasons to support a belief because it helps to track the subjects' commitment to the belief. Authorship of a belief is achieved by giving reasons for why the subject has the belief. Bortolotti observes that, of all the beliefs that are authored, only a subset will be authored in a "rational way", and therefore, agential rationality is more demanding than simple authorship:

> [W]hereas authorship requires that a subject is able to endorse the content of her beliefs *on the basis of what she takes to be her best reasons,* agential rationality requires that the subject is able to endorse the content of her beliefs *on the basis of reasons that are intersubjectively good*.(Bortolotti, 2010, p. 177)

Now, the question Bortolotti examines first is whether delusions are authored. This is a complex matter, because, in some cases, delusions are held with conviction, but the subjects do not offer any sort of reason, as it is the case in delusions of reference. In other cases, the reasons offered do not make sense to the interviewer but do seem to make sense to the patient, where things such as freckles can become a signal of

someone injecting poison for the delusional patient. There are more elaborate cases in which the report of the delusional patient is very detailed and some logical connections are made to explain (away) the strange phenomenon, but they are full of inconsistencies, or explanations with very low probabilities of being true, as is the case in the delusional patient in (Komiyama, 1989, p.15). Bortolotti states from this sort of interpretation that delusional people can give reasons to maintain their delusions, but those reasons are not intersubjectively good; and then, she poses the relevant question for this type of rationality: When does a bad reason surpass the line and cease to be a reason altogether?

Normal subjects also tend to confabulate irrational reasons when presented with some tasks, such as Nisbett and Wilson (1977), which case is very well summarized by Carruthers (2005, p. 142-143):

> Subjects in a shopping mall were presented with an array of four sets of items (e.g. pairs of socks, or pantyhose), and they were asked to choose one of them as a free sample. (All four sets of items were actually identical). Subjects displayed a marked tendency to select the item on the right-hand end of the display. Yet no one mentioned this when they were asked to explain why they had chosen as they did. Rather, subjects produced plainly confabulated explanations, such as that the item they had chosen was softer, it had appeared better made, or that it had a more attractive colour.

This example shows one of many cases of normal (everyday) confabulation where beliefs are not formed, nor justified, by intersubjectively good reasons. Bortolotti's conclusion is that normal beliefs also violate principles of agential rationality.

d) Bortolotti's doxastic conclusion

Bortolotti concludes that procedural, epistemic and agential rationality could be desirable for the concept of belief and belief ascription in general, but that we should not take these criteria as the ultimate constraint. She states that, given all the evidence about failures, our concept of belief should not be idealized.

Her positive account of delusion is that these irrational beliefs are a failure in self-knowledge, because, although they are authored (the subject endorses his delusion), they are not self-ascribed (they do not think there is something wrong with their own minds). Also, delusions have a strong impact on the self-knowledge of the person, because what they used to think about themselves and reality has changed. Therefore, the subject has gone through a dramatic change in his self-knowledge. In Bortolotti's opinion, this reinforces their status as beliefs.

The difference between delusions and non-delusional beliefs cannot lie in epistemic features. Her final thought is that delusions do not differ from normal beliefs in kind, just in degree:

> Delusions are on a *continuum* with irrational beliefs, and we find them towards the "very irrational" end of the spectrum, where the degree of rationality tracks both how much they deviate from norms of rationality for beliefs and how many norms of rationality they deviate from. (Bortolotti, 2012a, p.4)

e) Objections against Bortolotti's account

The first objection comes from Hohwy and Rajan (2011), who say that delusions are more like perceptual illusions, because:

1) Illusions and delusions are both unrevisable.
2) Both illusions and delusions can have varying levels of circumscription.
3) People with delusions have their competence intact but their reasoning performance is impaired due to deficient inputs.

Bortolotti responds that, first; delusions are not necessarily illusions because neuroscience has not been able to give a definite answer to how delusions are formed and maintained. Second, Bortolotti does not think that delusions are completely unrevisable, because there is evidence that cognitive probing and behavioural therapy work in some cases, and there are also instances where people have successfully overcome their delusions with appropriate therapy and medication. Third, Bortolotti argues that circumscription is variable, even in monothematic delusions. In some cases, patients tend to acknowledge there is something wrong with their delusion, and in some other cases, patients become violent and attack the object of their delusion. It is not

clear, in any case, how circumscription is supposed to work in illusions; not enough studies have been statistically made to measure how illusions affect beliefs and in what degree. Fourth, Bortolotti argues that, if Hohwy and Rajan assert that "people with delusions have their competence intact but their reasoning performance is impaired due to deficient inputs […]" (Bortolotti, 2012, p. 41), but then, their target is the two-factor theory of delusions. The latter is not necessary for Bortolotti's doxastic argument, she states, and although she has some sympathy for it, undermining it would not debunk the doxastic nature of delusions. Hohwy and Rajan make the following comment about doxasticism about delusions:

> [I]t is not a given that they are irrational beliefs, on a par with other irrational beliefs. Some delusions could very well be better understood as irrational perceptual inferences. (Bortolotti, 2012, p.43)

What they are trying to establish is that delusions and beliefs are not similar because the former are not caused by a reasoning competence failure. However, Bortolotti does not say that delusions are caused by a competence failure. She states that, for example, regarding the assimilation of counterevidence, delusions resist counterevidence to a greater extent than ordinary beliefs, but no different in kind to other irrational beliefs. In my opinion, Bortolotti successfully demonstrates that delusions are not illusions, but this is not a largely contested view in the literature.

Tumulty (2012) has a similar view to Bortolotti regarding the fact that there is a continuity between irrational beliefs and delusions. However, instead of ascribing to both the status of beliefs, Tumulty's proposal is that both are *not* beliefs: "when a subject violates a norm that shapes an important part of the dispositional profile for a belief, she fails fully to have the belief in question" (Bortolotti, 2012, p.50). Tumulty puts forward the idea that irrational states should not count as beliefs because they ought to have a regulative (not a descriptive) function, governed by the laws of rationality. Bortolotti counter-argues that in her account the regulative function is managed by self-narratives: in her positive account of delusions, Bortolotti observes that some delusions are not authored by the patient but he still integrates them into his personal story; she calls this phenomenon self-narratives.

> [W]hat seems puzzling about delusions is not their being irrational, but their badly fitting with an overall narrative that affects the interpretation of the subject's personal history […] (Bortolotti, 2010, p.257)

In my opinion, the role of self-narratives as a doxastic argument is not Bortolotti's strongest point. It is unclear how endorsement can contribute to a doxastic account: there are instances of acceptances that can be endorsed, but not believed. Bortolotti herself gives examples of such attitudes in her book: hypocrisy and false testimony are two of them. Therefore, in principle, Bortolotti would have to admit that Tumulty's option of delusions being equal to normal beliefs could potentially make both non-doxastic. At this point we have an underdetermination.

Tumulty and Schwitzgebel argue that delusions could instead be on a "sliding scale" in which thoughts could be classified as belief, or not, in degrees; this is useful for answering things such as if animals have beliefs when a simple yes or no answer will not be satisfactory:

> The sliding scale offers a sensible approach: the mental state in question can be considered more or less as a belief depending on the extent to which its features overlap with features of typical belief states. (Bortolotti, 2012, p.44)

The main disadvantage Bortolotti sees in this proposal is that it still endorses the rationality constraint on belief ascription, which she has taken to be completely undermined in normal beliefs throughout her book. Bortolotti also makes the clarification that, in order for something to be a belief, it should be endorsed as a belief. This leaves out some pathologies, such as thought insertion, but still, Bortolotti affirms, most delusions will satisfy that criterion.

In my opinion, these are damaging arguments against Bortolotti's account. Firstly, although she argues that the rationality constraint on beliefs is wrong, in her account she still maintains that all beliefs have a degree of irrationality. Conceding for the moment that there is no need for the rational constraint on beliefs, then it is not clear how Bortolotti would explain that rational, true beliefs exist in the first place. Also, she does not have something that replaces the role of rationality in her account. Therefore, although she says that she has undermined the rationality constraint, her account does

not destroy it. Secondly, Bortolotti argues that beliefs should be endorsed. However, that implies a circularity in her definition: in her account delusions are endorsed because they are beliefs in the first place. Thirdly, that the delusion of thought insertion is not explained is a damaging counterargument against her own answer. Therefore, Bortolotti cannot explain away any of Tumulty and Schwitzgebel's objections. In the end, Bortolotti's response shows that her account has not disposed of the notion of rationality in beliefs, and that her definition of endorsed beliefs is circular.

Schwitzgebel also has an alternative to the sliding scale, that is, that delusions need to have belief-like effects in order to be considered as a belief. Bortolotti states that this approach is supported partially by her own account, since delusions will still have some of many possible behavioural manifestations, such as endorsing their content or acting upon them. But the problem of a sliding scale is that it would have to, in principle, leave out delusions. But, since there is not enough evidence to consider delusions as non-beliefs, and delusions would have to be in the more irrational side, it is unknown what would actually lie outside the scale. Also, Bortolotti would say that her account has a more accurate description of the rationality of everyday beliefs and delusions than the sliding scale, therefore having a more positive (contrastive) view.

Nevertheless, Bortolotti's answer is not satisfactory. Although I do not consider that a sliding scale would solve the problems because it would inherit the problems from both antidoxastic and doxastic views, Bortolotti's counterarguments do not work in this case, precisely because her depiction of everyday beliefs is not about normal, functioning, true beliefs. Although true beliefs are not the only existing ones, these arguably are most part of everyday beliefs. Bortolotti seems to miss what it is important in them too by just placing them in a continuum.

Keith Frankish (2012) and Murphy (2012) have the concern that a single, folk-psychological definition of belief is unlikely to capture its diverse nature. Also, it seems to them that this folk-psychological definition of belief is very likely to be replaced by a more mature science of the mind, where the latter probably means some sort of neuroscientific definition. Another of Frankish and Murphy's options is that we could distinguish different types of beliefs and then ascribe one of these types to delusions.

Bortolotti concedes that it is tempting to do so because of the non-homogeneous nature of beliefs, but she argues that different types or levels will not capture the required essence either. The problem is that their positions "will not solve the problem of having

mental states that deviate from the standard" (Bortolotti, 2012, p. 49) because they are just descriptive and therefore will not provide any explanatory power; the latter cannot confirm or refute the doxastic nature of delusions. Also, Bortolotti accepts that her position is very minimalist:

> In the book I was attempting to offer reasons to divorce the intentionality of belief states from rationality, and to provide a minimal account of belief that could capture a general phenomenon without obscuring the obvious differences between types of belief. (Bortolotti, 2012, p.50)

But she states that no other position has been able to explain the phenomenon of delusions.

I think that Frankish and Murphy have a very important point: that the nature of belief is diverse in itself. To me, this is a salient feature of doxastic accounts of delusion. There is not a satisfactory account of beliefs that can encompass normal beliefs. Therefore, it is even more unlikely that a definition will encompass both normal beliefs and delusions. This is so because it is not the case that the dispute about the nature of beliefs is just verbal. It is a principled dispute because every account has their own, independent sources and assumptions. Therefore, that beliefs are not well defined is a very potent counterargument against the doxastic continuum of beliefs.

My criticism of Bortolotti's position will mainly focus on the explanation it offers regarding the content of delusions[12]. In particular, I will argue that her account fails to explain two key aspects regarding the content of delusions:

(i) The oddness of their content *per se*.

(ii) The oddness of the connection between the content and its justification – i.e. the reasons reported to support such content.

In general, Bortolotti's approach works by equating delusions and beliefs in two related ways: on the one hand, by inflating the irrationality of normal beliefs, and on the other, by deflating the irrationality of delusions. She draws two different conclusions from this. Firstly, she points out that the content of a particular delusion in a normal subject is not necessarily backed up by intersubjectively good reasons. Secondly, she argues that it is

[12] Although I address the problem of meaning and content in the next chapter, some of the objections are relevant specifically to Bortolotti's account.

not possible to have access to the background beliefs of a subject in order to assess whether the content makes epistemic sense to the latter.

We have to question now if this strategy is adequate. For instance, in the case of Cotard, the content is not just irrational, but also unthinkable. When asserted, the meaning of the utterance cannot be understood in terms of a possible situation of one's reality. One option is to understand the meaning of the assertion as a kind of metaphor; however, delusional patients deny that it is *meant* as a metaphor. The same situation can be observed in the case of thought insertion. What then is the odd belief expressed by subjects suffering Cotard, or in thought insertion delusions?

Bortolotti's dismissal of the notion of rationality for normal beliefs not only does not help in describing normal beliefs; more importantly, her strategy of equating the rationality in normal beliefs to rationality in delusions does not give a satisfactory answer to what is the content of delusion. It is not possible to just propose, as Jaspers (1913) does, that content is only circumstantial to delusions since the critical issue here is whether a subject really believes what he asserts. Let us consider the following: If it is not possible for an external viewer to point out what the actual belief of a patient is[13], then what is the 'belief' that Bortolotti is trying to defend? More specifically, what we are asking is: what element in Bortolotti's account is supposed to fulfil the role that rationality usually has? If we cannot resort to rationality in the usual way, then how can she explain what it means to ascribe beliefs to people?

Bortolotti argues that, by deflating the role of rationality, it becomes possible to ascribe the belief '*I am dead*'[14] to a particular subject with no problem. Her account suggests that '*I am dead*' is, in effect, a belief, although a very irrational one. Consider, however, that by taking the belief '*I am dead*' as no different in kind to others one is assuming an attitude of taking away the pathological, bizarre and strange character out of delusions. This, of course, cannot be done. In this sense, Bortolotti's account implies a blindness towards certain key features of delusions. In particular, this blindness towards their oddness will do no good to any adequate account. After all, how is an arbitrary observer to understand that the utterance '*I am dead*' reflects an objective belief of a subject if that utterance cannot be understood according to what the observer considers the usual

[13] Arguably, said belief can only be conveyed either by utterances in a specific language or as an implication of specific observable behaviours.

[14] By 'I am dead' I am specifically denoting the utterance by which the belief of (oneself) being dead is conveyed.

rules of 'rational' discourse'? In any case, the issue seems to have transformed into what do *we*, as observers, mean by those words when we use them to ascribe a belief to a subject? It seems that Bortolotti's account faces a serious objection. Arguably, an observer can rationally assess whether a subject holds a belief only by assessing the meaning of an utterance relative to some observable factors (i.e. behaviours); therefore, if we assume the possibility that subjects can hold irrational beliefs – such as those conveyed by the utterance '*I am dead*' – then either:

(i) We lose grasp on what it means for someone to hold a belief or

(ii) We lose grasp on what it means for an utterance to convey a rational belief – i.e. we lose grasp on which sentences can actually be said meaningfully

Therefore, Bortolotti's account of a continuum of beliefs is not adequate, in my opinion, to describe the nature of delusions.

IV. Delusions as anomalous beliefs

Marga Reimer, in her (2010) paper 'Only a Philosopher or a Madman: Impractical Delusions in Philosophy and Psychiatry', argues that psychiatric delusions are better described as a type of non-typical, anomalous belief that is ontologically similar to philosophical nihilistic beliefs. Reimer's motivation is clear: she wants to state that, although delusions do not seem to adhere to the characteristics of conventional, *typical* beliefs, that does not necessarily imply that their doxastic status should be undermined:

> My primary aim in this paper is simple: To defend the view that psychiatric delusions are genuine, if anomalous, beliefs. My defence is limited; I defend the view in question against the argument that psychiatric delusions are not beliefs because they do not "act like" beliefs. (Reimer, 2010, pp.316-317)

Reimer starts by recognizing that delusions do not have the characteristics usually attributed to normal beliefs. In delusions one often finds that the patient's behaviour is not always what would be expected from someone who has a given particular belief. However, still, in some cases the patient will act according to the delusion (Wessley et al., 1993). Sometimes patients might distance themselves from the delusion, as Currie (2000), Egan (2008), and Stone and Young (1997) remark: "The point is simply that such

patients do not *always* act as one would expect, on the assumption that they genuinely believe their delusions" (Reimer, 2010, p.316). The latter has led many philosophers to question the status of delusions *as beliefs*, and many non-doxastic stances have been motivated by the mentioned anomalies, such as Currie's (2000) delusions as imaginings, Egan's bimaginations (2008) and Berrios' (1991) definition of delusions as empty speech acts.

However, the latter consideration is not enough for Reimer to conclude that delusions are not beliefs. She maintains that the arguments put forward against construing psychiatric delusions as beliefs are similar in nature to considerations against a specific type of normal beliefs, that is, what she terms 'nihilistic' philosophical ones.

a) Psychiatric delusions have the same belief status as nihilistic philosophical doctrines

To Reimer, the most important problem in delusions is to try to explain why they do not influence the patient's behaviour as normal, typical beliefs do. In order to answer this, she sets out a counter-example in which a normal belief can act as delusions do. In particular, she claims that delusions are not the only instance of belief-like states that have a weak influence on the subject's behaviour: some particular philosophical doctrines, such as global scepticism, hard determinism and moral anti-realism have a similar role in the person's net of beliefs. Reimer states that subjects that endorse these philosophical views typically behave *as if* their everyday beliefs were not neutralized by their own philosophical stances, creating some sort of contradiction between their beliefs and the way they act upon them.

This prompts her to address the following question: "Why does the nihilist philosopher act (and apparently even think) contrary to the philosophical doctrines she devotes her career to defending?" (Reimer 2010, p.318), and Reimer has two main reasons to explain this:

First, she points out that these types of nihilistic philosophical doctrines are extremely difficult to maintain in ordinary everyday life. Philosophers, according to Reimer, simply act unreflectively, whilst their philosophical position may be a completely different one reflectively. That is, although a nihilistic philosopher might say that there is no knowledge we could acquire, she might still express things like "I knew that this cheese wouldn't last, it is mouldy". Reimer explains: "This rather striking difference is tied, no

doubt, to the fact that our unreflective beliefs are anchored in, and reinforced by, the realities of day-to-day living" (Reimer, 2010, p.318).

The second reason why nihilistic philosophies fail to influence behaviour is that people are aware of the negative consequences that their acting upon them would bring: "Whatever the results for the global skeptic, hard determinist, or moral anti-realist, they would not be good. The philosopher knows this, her nihilistic tendencies notwithstanding" (Reimer, 2010, p.318).

In general, Reimer's conclusion is that nihilistic beliefs, although considered as genuine beliefs, are "context sensitive". She says that this view has the advantage of saving the affirmation of the philosopher that she sincerely believes what she maintains whilst not always acting upon it.

On the other hand, we have the case of psychiatric delusions, which sometimes do not seem to influence the subjects' conduct as discussed in the previous chapter. Reimer seems to offer an interpretation in which their epistemic status looks similar:

> These considerations suggest that the role played by philosophical argument in the case of nihilistic doctrines might parallel the role played by anomalous perceptual experience in the case of (at least some) psychiatric delusions. When attending to the relevant sorts of philosophical arguments, the nihilist philosopher is convinced of the truth of her unintuitive doctrines; when attending to his anomalous perceptual experiences, the psychiatric patient is convinced of the truth of his outrageous beliefs. (Reimer, 2010, p.319)

For Reimer, this results in the phenomenon in which both the patient and the philosopher answer affirmatively when asked if they believe in their relevant beliefs, but they fail to act upon them in the absence of a stimulus. If this is so, are delusions like philosophical nihilistic beliefs? The answer is not straightforward because, as Reimer acknowledges, there are still three main characteristics that differentiate them.

b) Differences between philosophical doctrines and delusions

According to Reimer, the most important difference concerns the matter of evidence. Non-psychiatric subjects usually have "some sort of credible evidence for their beliefs.

(…) Such is the nature of the normal (non-delusional) subject's sensitivity to counter-evidence", whilst in delusions "the facts are thus interpreted so as to confirm the very delusion motivating the interpretation in the first place" (Reimer, 2010, p.320). Therefore, it is not the case that delusional patients ignore the relevant facts that undermine their beliefs; it is just that they accommodate them in a way that actually favours the delusion.

The second difference consists in the circumscribed role a delusion has in the subject's "web of beliefs". This is because delusional patients tend to leave unresolved links between their delusional beliefs and other relevant beliefs. In contrast, non-delusional subjects make constant adjustments and draw the obvious conclusions from their webs of belief without much problem, as happens in everyday life.

The third and last difference is the impact of delusions and ordinary beliefs on the subject. As I have discussed before, the delusional subject may show an abnormal lack of affect about what she is maintaining as a belief (although delusions can be very stressing in some cases too).

> [D]elusions lack the kind of holistic character that beliefs are supposed to have: they do not interact with perceptual input, other cognitive states or behaviour in the way that beliefs should. (Bayne and Pacherie, 2005, p.165)

Non-delusional subjects usually experience the expected emotions in the relevant circumstances.

So, should we conclude that these differences make delusions ontologically and epistemically different? Reimer thinks that these differences can be accommodated in order to preserve the claim that delusions are similar to philosophical doctrines.

To start with, Reimer argues that the first argument, which concerns the different attitude towards evidence, can be fixed by appealing to "common sense" philosophical doctrines. About these stances, Reimer observes that discussions such as G. E. Moore's (1939) epistemology, or the moral realism of McDowell (1978), have the characteristic of its truth being intuitively obvious *and resistant to contrary evidence or argument*:

> Such views are not of the sort that require evidence or argument; their truth is so patently obvious that even considering "evidence to the contrary" is

> unthinkable. And yet there are reasons, good reasons, to at least question these common sense beliefs. (Reimer, 2010, p.322)

Therefore, what delusions and these philosophical doctrines have in common is a kind of "intractability". The latter concept is closely related to the "subjectivity" motivating both beliefs, because "from your epistemically privileged point of view, your explanation of your thoughts and experiences might work perfectly fine" (Reimer, 2010, p.322). Reimer concludes that, if delusions are not genuine beliefs because of their intractability, these "common sense" doctrines should bear the same status.

About the second objection, in which delusional subjects do not integrate their thoughts into a web of beliefs, Reimer's opinion is very clear:

> At any rate, it seems that nihilistic philosophical doctrines do not always (although they might sometimes) influence one's web of ordinary beliefs. What such doctrines likely do influence is one's web of *philosophical beliefs.* (Reimer, 2010, p.324)

Reimer explains this lack of communication between beliefs as a Davidsonian (1982) *compartmentalization*, in which beliefs might be stored in different mental compartments that do not "talk" to each other, therefore generating some inconsistencies. The latter affects the epistemic context of beliefs and therefore ordinary subjects with nihilistic beliefs only practice them in certain relevant, philosophical contexts.

Reimer also considers the third objection where she points towards the tendency of delusions to not influence the subjects' emotions. She states that the endorsement of the nihilistic belief does not invariably lead to despair; however it might be the case that these doctrines could reinforce a depressive personality.

So, it seems that Reimer maintains that the main arguments that differentiate delusional beliefs from nihilistic doctrines can be surmounted: "These considerations, taken collectively, provide some support for the claim that both nihilistic and common sense philosophical beliefs are, like psychiatric delusions, unusual (atypical, non-paradigmatic) beliefs"(Reimer, 2010, p.324). In the next section I present Reimer's argument, the fallacy of the atypical, which tries to undermine antidoxastic objections.

c) The fallacy of the atypical

Reimer's second aim is to show that philosophers fall into the fallacy of thinking that if *x* fails to be a paradigmatic case of *y*, then *x* is not an instance of *y*. She explains that the fallacy consists in failing to see that *x* could be an *anomalous* case of *y*. So, following her arguments, both delusions and nihilistic philosophical beliefs are unusual (atypical, non-paradigmatic) beliefs. Both lack a direct influence on the subjects' behaviour, are intractable, and fail to influence their belief systems as a whole. She proposes an analogy in her (2010) paper, which argues delusions are anomalous beliefs, just as penguins are anomalous birds:

> A simple analogy should make the point vivid. Although penguins are birds, they are highly unusual birds. They combine the unusual features of other unusual birds. Like kiwis, they are unable to fly; like sawbills, they are able to swim underwater. They are, as it were, "doubly" unusual. But that does not mean that we should feel any pressure to conclude that penguins are "neither fish nor fowl." We can simply (and correctly) say that they are unusual birds. (Reimer, 2010, p.324)

Therefore, to Reimer, atypical beliefs are still beliefs, and this is the case in delusions. She argues that there is no point in considering an alternative view of delusions because the concept of belief is well equipped to deal with the main counterexamples against the belief status of delusions.

d) Reimer's final remarks

A sort of conciliating remark from Reimer is that she does not consider herself as someone that puts forward a strongly doxastic stance. Instead, she considers herself to be neutral towards it: "As to whether or not psychiatric delusions are genuine (if anomalous) beliefs, I remain officially neutral" (Reimer, 2010, p.317). She states that doxastic arguments, such as Bortolotti's (2005) and Bayne and Pacherie's (2005) are as compelling as Stephens and Graham's (2004). However, her arguments reinforce the doxastic stance, given that she thinks that the burden of proof "lies with those who deny

what is prima facie plausible: That certain psychiatric delusions, like certain philosophical doctrines, represent genuine, if anomalous, beliefs" (Reimer, 2010, p.317). Reimer explains why she is not arguing definitely in favour of a doxastic account: she thinks that the status of being a propositional attitude is not determined solely for being a typical or an anomalous belief. In my opinion, the important question she asks in her (2010) paper could be: "What are the necessary and sufficient conditions for a mental state's being a belief, if indeed there are any conditions?" Nonetheless, Reimer has her own, lighter criteria for asserting what a delusion is: for her, it is reasonable to assume that asserting *p* "constitutes prima facie evidence for supposing that the assenting agent genuinely believes that *p*, at the time of assent" (Reimer, 2010, p.325), therefore conceding that status to delusional thoughts. Further, her final remark is:

> [E]ven if psychiatric delusions turn out to be *sui generis*, that is no reason in and of itself to conclude that such delusions are not genuine beliefs (…); it should instead lead us to affirm the intuitive assumption that they are *anomalous* beliefs. (Reimer, 2010, p.326)

e) Criticisms

George Graham responds in his (2010) paper to Reimer that he does not wish to classify delusions as beliefs, however convincing Reimer's point might seem. To him, the question whether people are deluded has to do with the various thoughts, feelings and beliefs they have, and not with "attitudinal botany" (Graham, 2010, p.337). He maintains that "mad desires", as well as "Martian desires" might be different to "human desire" and therefore, he is not persuaded at all by Reimer's (2010) arguments: "I certainly respect her effort, but I am unmoved by it" (Graham, 2010, p.338). Graham's main objection consists in stating that philosophers actually distance themselves from their counter-intuitive beliefs:

> When a certain philosopher enters the conceptual world of his or her odd claims, they keep track of alternative convictions, including those of the conceptual world of the everyday. Some philosophers intentionally distance themselves from such everyday convictions and assume that the rules or standards of

> philosophy (of metaphysics in particular) are incommensurable with the rules or standards of everyday reasoning (…) The philosopher defends the truth of the bizarre claim from a perspective of everyday detachment—something that in a certain manner occurs in some cases of delusion (and helps to fuel suspicion that delusions are not beliefs). (Graham, 2010, p.338)

Graham then ends with a brief note, saying that it is the attitude (not the content) towards their own thoughts that determines who is delusional and who is a philosopher.

> Difference in belief is not as such what separates the nihilistic self-existence denier in metaphysics from the victim of the Cotard delusion. Each person's attitude toward the contents of their thoughts divides them, whether the thoughts qualify as beliefs or not. The first assumes a type of philosophical stance toward those contents. The second person is in the grips of the delusional stance. (Graham, 2010, p.339)

A different criticism comes from Tim Bayne (2010). He starts by affirming that not every antidoxastic position relies on what Reimer calls the "fallacy of the atypical". Although it is true that some philosophers do, most antidoxastic positions do not think that delusional beliefs are merely atypical.

> On my reading, the fundamental worry that motivates the anti-doxasticist view is the thought that delusions are anomalous in ways that are at odds with their putative status as beliefs. In a nutshell, the worry is that delusions fail to play the functional role that is essential to a state's being a belief. (Bayne, 2010, p.330)

So, according to Bayne, the argument of the fallacy does not apply to most of the antidoxastic stances.

Secondly, regarding the argument that delusions are similar to nihilistic philosophical beliefs, Bayne claims that there is a blunt epistemic difference between the deluded and the philosopher. The philosopher will still endorse her own nihilistic arguments even when acknowledging that she "knows" the location of the car keys. On the other hand,

a patient might have to be reminded about why he has to believe that he actually knows his own relatives.

> What fluctuates is not the philosopher's appreciation of the evidence for certain claims, but the costs of endorsing those claims. Philosophical contexts allow one to get away with assertions that one cannot get away with in everyday life. (Bayne, 2010, p.331)

A third observation in Bayne's (2010) paper is that, if Reimer's position works in a way in which delusions and nihilistic beliefs are replaced by ordinary beliefs and vice versa, Bayne's worry is that then the argument of compartmentalization is made redundant. This, Bayne says, can be saved by proposing that those beliefs can only be triggered when the "normal" beliefs are not happening.

Bayne argues that the question whether delusions are beliefs depends on each account's perspective. In this matter we have two options; the functionalist perspective and the normative one:

- In the functionalist perspective, according to Bayne, delusions do not look like beliefs, because functional roles are supposed to dictate what a paradigmatic role would be, due to the fact that delusions do not adhere to the functions that one could possibly define for the role of a standard belief. To Bayne, the difference between Reimer's position and Egan's (2008) is not too clear regarding functional roles, because both agree on the fact that the functional role of a delusion would be a state "in between" that is constantly mixing with other functional roles.
- On the normative perspective, the question arises of how committed to delusions one is. To Bayne, this is a central question because, seen from a normative point of view, both philosophers and deluded people can be charged with *inconsistency* and Reimer can be said to have some sort of say.

An important remark of Bayne's underlying assumptions can be seen in his own words, regarding his ideas of how the world is:

> Insofar as we embrace a realist conception of the world—the idea that there is a single way that the world is—we must reconcile our various truth-related commitments with each other. (Bayne, 2010, p.334)

But the actual paradox in the normative approach of the definition of beliefs is that this notion is not sufficiently uniform to encompass all the characteristics of delusion:

> Assuming this normative approach, do delusions qualify as beliefs or not? I doubt that delusions have the kind of unitary nature that would be needed in order for this question to have a determinate answer. Some delusions might be best understood as commitment-involving, in which case we can and should evaluate them with respect to the norms of belief. (...) Other delusions might be best thought of as a kind of imaginative charade, and not legitimately evaluated with respect to the norms of truth and rationality. The case for regarding a delusion as a doxastic state may differ from patient to patient and may even fluctuate for particular patients from one occasion to another. (Bayne, 2010, p.334)

So, following his objections to Reimer, the matter of whether delusions are beliefs or not is, at most, inconclusive. However, Reimer has been able to offer her point of view on these criticisms.

f) Reimer's response to criticisms

In her (2010a) reply, Reimer takes a stronger, more explicitly doxastic position: "I thus remain sympathetic to the idea that paradigm psychiatric delusions tend to be full-fledged beliefs and not simply opinions" (Reimer, 2010a, p.341). She defends her position by arguing first that the *content* of the philosophical and delusional belief is important (contrary to Graham's claim in his 2010 paper against Reimer). For example, although an Ungerian (Unger, 1979) nihilist might express the thought "I do not exist", the Cotard patient might assert that he is non-existent with the same words. But Reimer remarks that they do not mean the same thing at all. The philosopher arrived at this stance by rationalising, by deduction; the Cotard patient, due to his measurable lack of

affection, *feels* that he does not have a body. These differences, she argues, affect not just the stance taken, but the content too.

Next, Reimer attacks Bayne's position that philosophical beliefs are well formed and revised, whilst delusions are not. Reimer maintains that evidential claims of normal people and deluded people are very similar. When both the nihilist philosopher and the deluded person need to argue about something, such as a belief in an ankle injury, both of them will use evidence that will come from their subjective personal experience. The normal person will say she *felt* her ankle bend and crack, whilst the deluded person might say she *feels* like there are worms devouring her ankle. Both will resort to the same sort of evidence (their feelings), and both might cling to their self-diagnosis although confirmatory evidence could be provided by the doctors. The problem with the delusional person in the matter of evidence is that the deluded person has an anomalous, persistent experience that will not be (cannot be) discouraged by other evidence due to the overwhelming subjectivity that affects the patient. However, Reimer is not completely adhering to Maher's (1974) notion of delusions being just "rational responses" to "anomalous perceptual experiences". She endorses a two-factor model of delusions where the second factor is a rationality deficit:

> One might counter that the anomaly of perceptual experiences is not sufficient to explain the development and maintenance of psychiatric delusions because it would be possible to have the relevant sort of experiences without being psychiatrically deluded. It is the combination of anomalous perceptual experiences with profound deficits in rationality that leads to the formation of psychiatric delusions. (Reimer, 2010a, p.345)

At the end of her (2010a) Reimer makes an unexpected argument. She states that, whilst dreaming, one *believes* the content of the dream, whether it is within lucid dreaming or not. And then she says that dreamers are not considered irrational, no matter how strange the dream's content is. And certain studies, such as Scarone et al. (2008) suggest that the phenomenology of dreaming and schizophrenia might be similar. Therefore, her argument is:

> But if we do not call the dreamer irrational for not recognizing her dreams as such, why should we call the psychiatrically deluded patient irrational for not recognizing her "dreams" as such? (Reimer, 2010a, p.345)

However, she gives no further argument to support this claim, which is rather weak by appealing only to some possible (and not strongly confirmed) phenomenology.

In general, Reimer's response to the objections appeals to the underdetermination generated by the subjective experience of the patient. Her arguments do not dissipate the worries that Bayne (2010) and Graham (2010) advance against her account. Even more, in her responses she endorses more her doxasticism without giving enough reasons to support this decision. Her account is mostly inconclusive, with her arguments being, at the same level of the rival counterarguments. This is a consequence of her concept of belief being vague enough to encompass both normal, philosophical beliefs, and delusions.

In the end, Reimer's account is not satisfactory because its arguments fall under the same underdetermination as her rivals. Also, it lacks explanatory power regarding the bizarreness of delusion. Comparing delusions to a non-pathological way of thinking is not really helpful to explain the strange features of delusion.

V. Conclusion

Doxastic stances show that there is no definite answer to two fundamental questions about delusion: What are the fundamental constitutive elements of a belief? Is a delusion a belief? Regarding the first question, the different doxastic accounts are forced to produce a flexible concept of belief. It is clear that a conventional view of what constitutes a belief does not work with a controversial subject such as delusions. One can also observe that belief accounts are incompatible, to a certain extent, among doxastic accounts because their guiding assumptions are different. However, they all agree that the main assumptions to argue against are the antidoxastic ones. About the second question, whether delusions are beliefs, there is no agreement because the existing accounts cannot accommodate the bizarre or puzzling characteristics of delusion to an account that works for all the other non-delusional beliefs, even when the concept is more flexible.

Another important problem with doxasticism is that the meaning of the content of the delusion cannot be explained, especially in cases where the belief is bizarre, impossible or non-understandable. The Cotard delusion, where the patient expresses that he is dead, is the most representative of such accounts. Doxasticism needs to answer the problem of meaning because one cannot believe what one cannot understand. More accurately, the problem is for the external observer who tries to explain the delusion: how can the observer affirm that the patient believes something that the observer cannot understand?

In the end, the fundamental problem of doxasticism is that one starts with the idea that delusions do not seem like ordinary beliefs: they are obviously quite odd. One then argues that, for the reasons explained in this chapter, they are like beliefs. So that leaves the question, why do they seem unlike beliefs? The more one does to show that they are akin to beliefs, the harder it will be to account for their strangeness as beliefs.

Comparing doxastic accounts with antidoxastic accounts might give the impression that they disagree on the important assumptions at first glance. However, they have many assumptions in common. One of these common assumptions is that the content of the delusion is important. Doxastic accounts argue that, for example, being chased by the CIA is not impossible (although it is rather improbable), and therefore, it should count as a belief (perhaps an unusual belief, as Reimer says). Antidoxastic accounts argue that the content itself is sufficiently strange, and the evidence given about the matter is not good enough, therefore, it cannot be a belief because beliefs are holistic and coherent with the subject's other beliefs. So, both doxastic and antidoxastic accounts agree on the fact that content is an important feature regarding delusions. Some other shared assumptions are:

*They argue that delusion is some sort of propositional attitude.

*Factors such as conviction, behaviour, and epistemic support will dictate what sort of propositional attitude a delusion should be .

*Contrary to what Jaspers says, whether a delusion is a belief or not, constitutes an important part of answering the question: What are delusions?

Therefore, their disagreement is only about one fundamental claim, but they agree about many other aspects in which the ontology of delusion should be seen. However, fundamental agreement cannot be reached from any of the stances because it is a compromise between being too ad-hoc and still missing the point (antidoxastic

accounts), or trying to force delusions into a familiar concept where they do not fit too well (doxastic accounts).

Given that there is no satisfactory answer from either doxastic or antidoxastic accounts about the nature of delusions I will analyse a radically different approach, that is, modern phenomenological stances that consider different assumptions to construct a theory of the ontology of delusions.

CHAPTER IV- PHENOMENOLOGICAL APPROACH TO DELUSIONS

I. Introduction

An important branch exploring the nature of delusions is that established by contemporary phenomenological accounts. Regarding the debate about the nature of delusions, the work of Jaspers (1913) is preferred over that of other phenomenologists such as Heidegger and Husserl. However, although contemporary phenomenological accounts are greatly influenced by Jaspers' work, they provide a different, fresh approach that may question the latter's assumptions.

It is worth analysing various examples of new phenomenological approaches so as to provide contrasts to the mainly analytical traditions expressed in the debates between doxasticism and antidoxasticism discussed in the previous chapters. After all, assumptions questioning whether delusions are beliefs or not are not the only ones worth taking a stance on.

Phenomenological approaches differ greatly from the ones in the previous chapters. Although the same empirical data has to be explained -such as, why there exists compartmentalization in patients with Capgras- the importance given to the facts tends to differ. For example, in phenomenological accounts the fundamental question is: What is the nature of the patient's experience? 'Experience' is not, however, meant as a definite set of measurable facts, such as what is the patient's heart rate or what are his dopamine levels. Rather, it is the opposite view, in that the patient's subjective experience is the focus of the account. This emphasis on subjectivity makes phenomenological accounts attractive to patients as a description of their illnesses; its richness is worth analysing as a possible answer to the nature of delusions.

However, as I argued in previous chapters, phenomenological accounts, too, face systematic objections. In this chapter I will analyse two examples: Shaun Gallagher's account of alternative realities and Matthew Ratcliffe's more explicitly phenomenological approach. In both cases I consider their main assumptions and assess the balance of the virtues and objections of the two accounts. Although there are many other phenomenological accounts about delusion, such as Gipps et al. (2004), Stanghellini (2008), Kapusta (2010), and Mishara et al. (2013), amongst others, I chose to include Ratcliffe's and Gallagher's account in this chapter because they are very

influential within the debate and their accounts portray an interesting contrast to more analytic stances.

II. Shaun Gallagher's account of Delusional Realities

In his (2009) chapter, 'Delusional realities', Shaun Gallagher develops the view of delusions as 'alternative realities'. His account does not aim to provide a causal explanation of delusions; instead, he intends to propose an accurate description that would provide a framework for any such future explanation. Gallagher suggests that his view is able to encompass some characteristics of delusion that are not always considered, such as affect, social and environmental factors.

Gallagher differentiates his approach from bottom-up, top-down and other hybrid approaches by maintaining that they share an *internalistic bias* in which organic malfunctions, such as abnormal brain processes, play a causal role. A top-down versus bottom-up distinction is provided by Campbell:

> On what I will call a rationalist approach to delusion, delusion is a matter of top-down disturbance in some fundamental beliefs of the subject, which may consequently affect experiences and actions. On an empiricist approach, in contrast, delusion is a rational response to highly unusual experiences that the subject has, perhaps as a result of organic damage. (Campbell, 2001, p.89)

Gallagher suggests that neither of those theoretical frameworks is sufficient to explain delusions because they ignore the embodied, social, affective, and environmental factors he favours. He observes the following "Is it possible to take a more externalist approach that can answer some of the unanswered questions about delusions?" (Gallagher, 2009, p.246). His preliminary answer is that this approach would have to take in account the fact that cognition is generated in a system that includes the brain-body system and the environment; not only sub-personal level explanations. Therefore, Gallagher proposes a hybrid account that takes in account such factors, and is compatible with his multiple reality hypothesis. First, I will focus on Gallagher's criticisms of bottom up and top-down accounts; after that I will describe his hybrid account.

a) Against top-down and bottom up accounts

Gallagher points out some of the dilemmas associated with top-down; one of the difficulties being *the problem of specificity* (Gallagher, 2009):

> [I]t is not clear why the subject's delusions are selectively about certain topics and not others – that is, why the subject is delusional in regard to some topics but not delusional in regard to everything he believes, or why some actions or thoughts are considered alien, but not others. (p.250)

Another related problem is that top-down accounts do not provide a clear explanation of how an organic malfunction could cause a *failure in cognitive mechanisms*. Campbell's account only provides a certain hint that organic malfunction could be the cause of delusions, or, in the case of Graham and Stephen, the fault could be an abnormal introspection mechanism. However, according to Gallagher (Cfr. 2009, p.250), it is far from clear how this mechanism works.

Gallagher also points out the paradox of *double-bookkeeping*, described by Sass (1994, p.21):

> [M]any schizophrenics who seem to be profoundly preoccupied with their delusions, and who cannot be swayed from belief in them, nevertheless treat these same beliefs with what seems a certain distance or irony […] A related feature of schizophrenic patients is what has been called their 'double bookkeeping' […] A patient who claims that the doctors and nurses are trying to torture and poison her may nevertheless happily consume the food they give her; a patient who asserts that the people around him are phantoms or automatons still interacts with them as if they were real.

Another one of his main objections to top-down models is that they fail to explain the specific content of the delusion. In monothematic delusions, such as Capgras, the content is always related to a loved one being replaced by an impostor and the delusional content does not vary. Accounts such as Campbell's cannot explain why one explanation is preferred among all the other possible explanations for an anomalous phenomenon.

Bottom up-accounts, such as Gold and Hohwy (2000), and Sass and Parnas (2001), suggest that the alien nature of the delusion comes from experience itself, like some sort of "experiential irrationality".

> We claim […] that the source of thought insertion and related delusions is the experience itself of the schizophrenic subject, and, in particular, its alien quality. The elaboration of the delusion in hypotheses and ancillary beliefs should be understood to be derivative from, or secondary to, this experience. […] The alien quality of the delusional experience is part of its content, and it is the content of experience that is the locus of the delusion and thus of the irrationality. At least some delusions, therefore, are best explained as disorders of experience rather than disorders of belief, desire, or reasoning. (Gold and Hohwy, 2000, p. 160)

Gallagher (2009) points out that there are several problems with the non-hybrid, bottom-up accounts. The first is that, although delusions consist of an immediate phenomenological experience, this may not be enough to explain all aspects of delusion. That is, the pathogenesis cannot explain, on its own, the bizarre narrative encountered in delusions. For example, in the case of Capgras, the specific lack of emotion towards one's partner does not necessarily point to a substitution hypothesis.

The second problem that Gallagher addresses is that it might still be possible to have some sort of top-down confabulation in delusions. According to the author it is not possible to reach such bizarre conclusions without some salience effect, in which the patient finds that he is more sensitive to otherwise innocuous stimuli, combined with confabulation.

Therefore, Gallagher suggests that a broader phenomenological account of delusion, combined with a hybrid approach, is a more appropriate solution to those problems.

b) Gallagher's hybrid account

Gallagher (2004a) proposed a hybrid bottom-up model which portrays experiences as immediate, non-inferential and non-introspective. His model makes the following distinctions (Cfr. Gallagher, 2009, p.251):

1) There is a first-order experience- a lived-through experience in the world.
2) There is a higher order cognition- that is, a reflective experience.
3) There are non-conscious processes -, sub-personal level, brain processes.

According to this account, self-agency problems are caused at a neurological level; by self-agency problems, the author refers to cases in schizophrenia in which the patient feels that he is being controlled by an external force. These cases are sometimes considered as cognitive errors, and Gallagher (2009) concurs. He suggests that in those cases the primary cause of delusion is brain pathology. Although the idea of cognitive errors is also present in top-down accounts, cognitive errors of this sort "are so deeply entrenched in perception, and so automatic, that they are an integral part of our *immediate, phenomenological experience"* (Gallagher, 2009, p.252). Consequently, their influence is pre-cognitive and non-inferential. Salience effects (such as a heightened sense of consciousness) are a possible instantiation.

According to Gallagher, these salience effects could shed light on the problem of specificity:

> Salience effects may also help to explain the problem of specificity – for example, the fact that in the case of thought insertion, specific kinds of thought contents, but not all kinds, appear to be thought inserted (Gallagher, 2004b) – if specific events, things, or persons trigger the dopamine release. In effect, schizophrenics may experience a certain semantic or content consistency amid the inconsistency of their delusions.
>
> Although it is difficult to explain the problem of specificity in purely sub-personal terms (e.g., Frith's 1992 suggestion about the dysfunction of a sub-personal comparator), salience effects at the first-order level of experience may contribute to an explanation. (Gallagher, 2009, p.253)

c) Gallagher's multiple realities hypothesis

Gallagher proposes a model of delusion that takes into account an embodied, phenomenological view of cognition, similar to Jaspers' view. Recall that Jaspers rejects the view of delusions as false beliefs as a partial account:

> To say simply that a delusion is a mistaken idea which is firmly held by the patient and which cannot be corrected gives only a superficial and incorrect answer to the problem [. . .] [A]ll experience of reality [. . .] has a root in the practice of living [. . .] Delusion proper [. . .] implies a transformation in our total awareness of reality. (Jaspers 1997, pp. 93-94)

For Gallagher, the subject is more than an intellectual creature who has beliefs and acts upon them. The subject, instead, is an embodied being, who acts as an agent and is physically, affectively, and socially situated in the world. This motivates him to suggest considering delusion as multiple realities.

In general, outside the context of delusion, multiple realities can be thought of as those realities in which one engages whilst dreaming, seeing a movie, or a theatre play, which are some sort of sub-universes or "finite provinces of meaning" (Gallagher, 2009, p.255). The various realities are not necessarily commensurable, and actions can be different in those different realities. With respect to delusional realities:

> It seems quite possible that one can enter into a delusional reality just as one can enter into a dream reality, or a fictional reality, or a virtual reality. Like other multiple realities, some delusional realities are ones that are more or less cut off from one's everyday reality; ones that are incommensurable with the normal rules of reason that govern one's everyday normal lifeworld, and ones that offer a different set of affordances. (Gallagher, 2009, pp.255-256)

A characteristic of the multiple realities account is that it must be firmly sustained, although this may vary in degree. Another characteristic is that one may slip in and out of the alternative reality. Gallagher also points out that some patients report a feeling of strangeness in their delusion (Stone and Young, 1997, p.337), consistent with the strangeness of being in a delusional reality. Also, in extreme cases, the delusional reality might overcome everyday reality.

Gallagher believes that doxasticism does not reflect accurately the dimension of an alternative reality: "To consider a delusion to be merely a belief is, as Jaspers suggested, to abstract it from something much richer – something that the delusional subject experiences and lives through." (Gallagher, 2009, pp. 257). However, Gallagher explains

that the lack of doxasticism does not necessarily mean the suspension of belief. The subject might still integrate several beliefs, even framework beliefs, about the delusional reality. These beliefs might or might not be false according to the everyday reality. The actual problem with delusions is "the failure to suspend belief in the ontological actuality of the delusional reality" (Gallagher, 2009, p.257). My interpretation of the latter is that the problem is actually to be able to migrate from the delusional reality to the everyday world, like in a dream. Therefore, for Gallagher, delusion as an alternative reality is not about having a set of false, possibly odd, beliefs, but about having a broader, phenomenological, experience of a reality.

Structuring his account of how subjects experience alternative realities, Gallagher draws on Heidegger's description of "being-in-the-world". This has a three-fold existential structure. The first element is the idea of attunement, in which delusional subjects have a fundamental kind of emotional stance, or mood, which differentiates the everyday world from the delusional one. The second element is understanding. This is manifested in a process which gives meaning to things and events. In delusional subjects, sometimes this understanding is impaired, as in psychotic depression, but sometimes it might be enhanced, like in *angst*. The third and last element concerns expression, and it is similar to the Wittgensteinian idea of framework propositions, which express some sort of pre-cognitive being-in-the-world. I will return to this idea in describing John Campbell's work in the next chapter. Gallagher comments: "the subject expresses the logic of the delusional reality from the perspective of that reality" (Gallagher, 2009, p.259). According to Gallagher, Heidegger's account of being-in-the-world helps to correct a mistaken emphasis of thinking about delusions as a set of false beliefs about the objective world; to him, delusions are mainly something experiential.

d) Solving problems with the multiple realities hypothesis

As Gallagher maintains, his theory is designed to be more comprehensive than pure bottom up sub-personal level explanations or top-down explanations.

The first problem Gallagher intends to solve is the *double book-keeping paradox*. In the latter, patients are not fully engaged with their delusions, acting sometimes in a way that it is not consistent with them. Gallagher argues that double book-keeping patients might be able to shift from delusional reality to everyday reality, and therefore, they

might be able to realize the strangeness of their delusion. In order for this hypothesis to be useful, Gallagher makes a suggestion for possible future research:

> What needs to be investigated here is the frequency and degree to which a patient can shift between multiple realities as they move in and out of delusional states, the nature of the transitions, whether this shifting is more frequent in prodromal cases, or in cases of partial remission, and so on. (Gallagher, 2009, p.260)

The second problem Gallagher addresses is a possible explanation for the complex nature of delusional narratives. The proposed theory is that, some sub-personal level experience that starts like a strange feeling, can only develop its complexity given a different experiential framework, that is, if the feeling finds support in an alternative reality.

The third problem considered by Gallagher is the *problem of specificity*, that is, the problem of why delusions just target some topics and not all the ideas that delusional patients have. According to him, a bottom up explanation is not adequate because brain dysfunctions are not something that can be turned on and off without some other factor being involved: "If the other factor is simply another sub-personal mechanism, then delusions should be randomly about anything, rather than, with a certain degree of consistency, about specific topics or things" (Gallagher, 2009, p.261). Gallagher's solution consists in arguing that the multiple realities are motivated by social and environmental factors encountered in everyday reality. Further, emotional reactions to some specific experiences may trigger the sub-personal dysfunction (Cfr. Freeman and Garety, 2003). The way embodiment works in Gallagher's theory is reflected in the phenomenological experience:

> The force of the emotion, if not consciously experienced, might be cashed out in terms of embodiment – in measurable autonomic processes, for example, and may show up phenomenologically as a certain way of being-in-the-world. (Gallagher, 2009, p. 262)

The fourth problem Gallagher intends to solve is the fact that there are shared types of delusions. He suggests that certain shared features of the environment could have an impact on the persons' culture. His answer is that, just as there are certain universal themes in literature and in games, one should expect certain particularities to be shared in the delusional realities as well.

In conclusion, Gallagher's proposal is that delusions go beyond being complex belief systems, therefore being more global than the level offered by doxastic accounts.

e) Objections to Gallagher

Gallagher aims to answer the question "Is it possible to take a more externalist approach that can answer some of the unanswered questions about delusions?" (2009, p.246). In order to consider this matter, one has to ask whether his arguments in favour of environmental factors and embodiment are good enough to offer an alternative to approaches that take into account mainly sub-personal level factors. To do this I will assess Gallagher in three categories: Are his arguments explanatory, justified, and intrinsically valuable? I argue that although his account is intrinsically valuable, his arguments lack explanatory and justificatory power.

I will first focus on the explanatory value of his arguments. Gallagher claims to have solved four problems concerning delusions that neither bottom–up nor top-down approaches can handle. He says that he has explained away the problems of double book-keeping, specificity, complex nature and shared types of delusion by proposing them to be alternative realities to the everyday reality. It is true that other theories about delusion cannot explain these bizarre characteristics of delusions, but I would like to question the true explanatory power of Gallagher's claim. In first place, Gallagher's explanation is vague in the sense that his solution could apply to many other problems, such as the problem of the great diversity of normal thoughts. One could say that normal thoughts have certain failures or bias because one can engage in a multiple reality where that specific bias is part of the premises of that particular world. Or, Gallagher's proposal could also be applied to other paradigms, such as why science is so diverse in its methods. The answer would be that scientists engage in multiple realities while engaging in their scientific work, and that these realities do not have to necessarily overlap, so that anybody could have his own methodology according to his own point of

view. Gallagher's proposal also leads to a sort of solipsism, where everybody has his own world, and this can change at any time; in the case of delusional patients this world changes without them having any awareness of it. Nothing seems to rule out the possibility that everybody, not only delusional patients, can have their own sub-world and that this world may change with their own rules. Therefore, given this explanation, nobody can know when one is engaged into what Gallagher calls the "everyday" world.
The second criticism is whether Gallagher's proposal is justified. It is clear that Gallagher is not very concerned about justifying his explanation of delusion (by contrast with serving as a framework for describing delusions), but he still engages in the process of justification in the first part of his chapter by comparing his account to other explanations of delusion: Gallagher states that other accounts such as bottom-up, top-down and some hybrid accounts do not encompass all the underlying factors. Although Gallagher's account is certainly more complete than other explanations in the sense of taking into account environmental, cultural and embodiment factors, the importance of these factors is not sufficiently justified. Gallagher is only pointing out that such factors *may* be the reason for the bizarre characteristics of delusion, but there is not a strong justification for these as there is in, for example, one factor versus two factor accounts of delusion. If, for example, there is an environmental cause for delusion, Gallagher does not support this with any study. Justification for these assertions is necessary, even if Gallagher merely intends his approach to be a mere framework where other theories may develop.
Gallagher puts emphasis on the fact that doxastic theories are insufficient to provide a complete account of delusions. In his view delusions are neither doxastic, nor non-doxastic, because neither one of them reflects the phenomenological, person-level nature of delusions. This is one of the most interesting parts of his work because he proposes an explicit alternative to the debate. He thinks that delusions might or might not be a belief, or a set of them, and instead, they can be captured within a different approach altogether. Gallagher argues that looking at the whole experience of delusion, taking into account externalist factors is a way of saying that beliefs, on their own, cannot define the delusional experience. This may be a valid objection given that environmental factors do seem to be involved in the development of delusions in patients, but such claim does not support the opinion that occupying a delusional reality is not a matter of having beliefs.

Finally, I would like to analyse the intrinsic nature of this view as a framework for describing delusions. I think that this framework is worth taking into account on its own merits because it has a unique insight into how being delusional *feels.* Most approaches in fact deny this by saying that delusions are "ununderstandable" (Jaspers, 1913). Gallagher's is a very intuitive description of delusions, and it is, possibly, a description that many patients could endorse. Whilst the idea of occupying an alternative reality seems a powerful metaphor, Gallagher does surprisingly little to clarify what he means by it. After all, there remain significant disanalogies between immersing oneself in the fictional world of a book and experiencing a delusion. Delusions are not very like fictional alternative realities.

Summarising, Gallagher's account is intrinsically valuable because it is empathic to the patient's experience. However, it is not well justified because, whether it is a candidate for empirical support or merely a possible framework, is too vague. Gallagher's view also makes the explanatory status seem less important, such as clarifying what is pathological about them. Finally, it is rather ambiguous in its central idea of an alternative reality which can be applied to a range of different, unrelated phenomena.

III. Matthew Ratcliffe's concept of delusional atmosphere and radical empathy

I will now turn to another account which draws more explicitly on the phenomenological tradition. Matthew Ratcliffe, in his (2010) paper, suggests that delusions can be understandable, contrary to what Jaspers maintains in his *General Psychopathology* (1997). His way of reaching understanding of the patients is an encompassing concept which he calls "radical empathy". He also suggests that, in order to understand the nature of the delusion, it is important to analyse the experience of the patient through the concept of delusional atmosphere, which I discussed in my chapter on Jaspers. Both concepts are interrelated, and I will explain this in the following sections. As a preamble, Ratcliffe briefly analyses the state of some of the contemporary analytic accounts of delusion and their main assumptions.

a) Doxastic assumptions

Ratcliffe's (2010) paper "Delusional Atmosphere and Delusional Belief" maintains that non-phenomenological accounts have shared assumptions that permeate the fundamental concepts of the debate about delusions. For example, Stone and Young's (1997), and Davies' (2001) two-factor empiricist accounts share with Campbell's (2001) rationalist account "a commitment to explaining delusions in terms of causal relationships between perceptual and belief contents" (Rattcliffe, 2013, p.229). These accounts imply that a relationship between those two is a common preconception of analytical accounts. But, according to Ratcliffe, such assumptions might come from a misguided preconception of belief. He points out that, for Jaspers, the definition of delusion as a mistaken fixed idea does not shed light on the nature of delusions. The problem, for Jaspers, requires an empathetic attitude in order to grasp the phenomenon; the possibility of clarification comes only through phenomenological dialogue. However, Ratcliffe concedes that the adoption of a phenomenological point of view does not imply the rejection of other accounts of delusions, since the two are complementary.

Ratcliffe acknowledges that contemporary literature on delusions still provides accounts that propose a doxastic analysis of delusions: "Others, who continue to maintain that delusions are beliefs, have stressed the need for more permissive or nuanced conceptions of belief (Bayne and Pacherie, 2005: Bortolotti, 2005)". Ratcliffe agrees with Jaspers, however, for whom the question of doxasticism in delusions would not provide us with an adequate answer to the essential nature of delusions:

> […] conceiving of delusions principally in terms of beliefs is not so much incorrect as beside the point; there is a failure to engage with what is most fundamental to them. (Ratcliffe, 2010, p.230)

Ratcliffe recognises that there are alternatives to doxasticism; for instance, Berrios (1991) suggests that delusions are empty speech acts, and Currie (2001) proposes that delusions are not beliefs but mistaken imaginings. However, he suggests that these alternatives are not answering the relevant question either: what Jaspers would say is

that we are still missing an adequate understanding of the nature and role of *perceptual changes* that are mainly responsible for the belief.

What is important to Ratcliffe, therefore, is the experience itself, and this is closely related to the sense of reality: "the experience within which delusion takes place is that of experiencing and thinking that something is real" (Jaspers, 1913, p. 93). According to Ratcliffe, this sense of reality involves a feeling of what it is to be real or unreal, and, significantly, a sense of reality is not exclusive of attitudes of belief, because it should be *integral to the perceptual experience*. For Ratcliffe other kinds of intentional states such as remembering, imagining, heavily depend upon the experienced sense of reality. Therefore, when that sense of reality is altered the form of these intentional states changes as well because delusional atmosphere is an all-encompassing change in the shape of experience and thought.

b) Delusional atmosphere and radical empathy

Ratcliffe considers that it is possible to understand delusions, in a different sense than Jaspers, the *delusional atmosphere*. Delusional atmosphere is a change in the horizon of possibilities in an experience; this horizon is formed by the various kinds of possibilities from an object's structured system. These possibilities are practical and perceptual, "some of which take the form of potential activities and others as potential happenings" (Ratcliffe, 2013, p.237). For example, a cup has the possibilities of being blue, being able to be held, being able to contain coffee, etc. It is important that, to the subject, these possibilities are *significant*, and are inextricable from our embodied phenomenology.

According to Ratcliffe, delusional atmosphere demands a distinctive form of empathy. In non-delusional contexts we tend to empathize with people because we take for granted that both sides inhabit a common modal space, and we assume that other people, the world, and the objects, have the same kinds of possibilities that we do. Because delusions involve changes in this structure a different sort of empathy is needed in order to achieve phenomenological understanding of them. Ratcliffe calls this new sort of understanding *radical empathy*.

Radical empathy, according to Ratcliffe (2014), does not consist in *simulating* another person's experience, nor is it to *experience* the sensation oneself. Empathy, in depression and in delusions, is a perception-like exploration through one's own limits

and points of view that develops through certain adequate styles of interpersonal interaction. This is an immediate experience as well, and not a process of inference. It is important to note that having an experience is not enough to have empathy.

> The kind of care or concern that is needed for empathy is to be identified with the attentiveness and openness to the other person that is partly constitutive of those process (…) When we enter into something we do not depart from our own perspective to better understand it. (Ratcliffe, 2014, p.278)

Delusional atmosphere has three main features according to Ratcliffe:

a) The first is that there can be a change in the patient's sense of inhabiting a shared world, which involves a certain loss of interpersonal possibilities from the experience. In this approach the line between the patients' experience and how the world is becomes blurred, therefore resulting in a quasi-solipsistic predicament. This view is endorsed by Louis Sass (1994) whom I will consider in the next chapter.
b) The second feature (also described by Sass) is the characteristic of having a practical disengagement or passivity. The world does not offer any more practical possibilities and therefore the patient becomes a merely voyeuristic observer.
c) The third aspect, central to many descriptions, including Jaspers', is an "all-enveloping sense of novelty or surpise (…) Hence everything seems odd; there is a pervasive feeling of its being somehow contrary to how it should be." (Ratcliffe, 2013, p.239)

To Ratcliffe the doxastic status of delusion cannot be judged because the limits of the notion of belief are not well defined:

> So the first thing to be decided is whether "belief" applies to contents that one accepts of an already given world and also changes in the form of experience, or just the former. A further complication is that a change in the overall shape of the experience not only disposes one towards certain beliefs; it also alters the form of belief. (Ratcliffe, 2013, p.240)

Ratcliffe points out how Jaspers' own concept of belief involves a double reality:

> Belief in reality can range through all degrees, from a mere play with possibilities via a double reality- the empirical and the delusional-to unequivocal attitudes in which the delusional content reigns as the sole and absolute reality. (Jaspers, 1913, p.106)

When Jaspers uses the notion of a "double reality", he points towards the phenomenon of double book-keeping, where the person seems to have at the same time two contradictory beliefs. Jaspers' account suggests that the question of whether or not delusions are beliefs is secondary because it has been formulated in a too simplistic way. "In so doing, it illustrates how engagement with psychopathology has the potential to both complicate and illuminate philosophers' conception of belief" (Ratcliffe, 2013, p.241).

At the end of his (2013) paper, Ratcliffe feels tempted to offer an approach that is based on neurocognitive findings in the section that he titles "Non-phenomenological understanding". He states that there is a neurobiological approach that is consistent with Jaspers' "predictive coding" approach. This finding suggests that the brain is concerned with successful prediction.

> At a non-phenomenological or "subpersonal" level, the brain is sensitive to mismatches between what is expected and what actually occurs. In the event of a mismatch, an error signal is generated, which triggers processes that work to reduce or eliminate future mismatches. It has been suggested that at least some delusions arise due to a global misalignment between predictions and error signals. (Corlett et al, 2009, p.1)

Therefore, to the brain, everything is presented as different from what was expected, despite the genuine correlation between the expected and the perceived. The phenomenological correlate of this is that everything should look more salient due to being consistent with the delusional atmosphere. Corlett et al (2010, p.348) maintain that "delusion represents an explanatory mechanism, an attempt to impose order on a

disordered perceptual and cognitive world". As I discuss in the next section, Ratcliffe uses the concepts of delusional atmosphere and radical empathy to apply his phenomenological account of delusion to cases of Capgras.

c) Horizon and the Capgras delusion

Ratcliffe uses the concept of "horizon" and "phenomenological affect" to explain the Capgras delusion. Since Capgras is known to involve some sort of anomaly of affect, it is important for Ratcliffe to clarify its meaning. Some authors treat the concept of "affect" as an unconscious physiological response, such as Davies et al. (2001, p.140), the absence of which results in an altered experience. Others treat it as phenomenologically accessible feelings, such as Stone and Young (1997, p.327). To Ratcliffe it is important to make the following distinction, that the absence of physiological affect is not necessarily lack of feeling, and does not imply the absence of feeling. Furthermore, patients do not report a *loss of the feeling of familiarity* but a *feeling of unfamiliarity*, which is phenomenologically different.

This brings out the issue that the feeling of strangeness is an important part of the experience. Ratcliffe suggests that *affect* is an *existential feeling*, meaning that it will involve both bodily feelings and ways of finding oneself in a world at the same time.

Ratcliffe's question can be formulated as "How could a 'feeling of unfamiliarity' contribute to the belief 'this person is not my spouse'?" His answer is that affect is not just a way of having bodily states; it shapes experiences and contributes to the interpersonal world thus experienced. The nature of its contribution, Ratcliffe says, is that in which a feeling of unfamiliarity can simply *be* the experiential content "this person is not my spouse but an impostor".

In contrast to Campbell's (2001) account (discussed in the next chapter) where the proposition ''that [currently perceived] woman is not that [remembered] woman", Ratcliffe considers that focusing only on the propositional content is an impoverished conception of experience because affect, or the lack of it, cannot be distinguished from content. He remarks that the lack of affect, on its own, cannot give rise to a delusional idea, because, for example, in decaying marriages, the person may become estranged from a close relationship without endorsing the belief "this perceived person is not that remembered person". But then, how is it possible to explain the following paradox?

> The patients often claim to be able to tell the difference between the original and the impostor, although they are unable to explain convincingly how they do this. (Stone and Young 1997, p.334)

> The essential paradox is that patients with Capgras delusion simultaneously recognize a face and, at the same time, deny its authenticity. (Ellis and Lewis, 2001, p.149)

For Ratcliffe, in order to solve the paradox we have to explain why the *actual* is not all that there is to the experiential content. This is solved by proposing the notion of *horizons*. In these horizons, the experience of something is not only constituted by what actually appears, but it is about the structured relationships between the actual appearances and their possibilities.

The notion of horizon, applied to affect and a sense of experience results in the following view:

> Stripped of distinctively personal possibilities, someone would not appear as a person at all but as an impersonal object that looked like a person, a model, imitation or fake, waxwork, manikin or mechanism. In one sense, the experienced entity would look exactly like a person. It would still incorporate the possibility of perceptually accessing its hidden aspects and of all those aspects being consistent with the characteristics of a person. But, at the same time, the entity would not offer up any of the possibilities constitutive of the personal. (Ratcliffe, 2008, p.212)

This motivates the content of the experience “this is an impostor” rather than the content “this person is not my spouse”. Therefore, it is this artificiality that characterises the experience due to a loss in the personal possibilities in the horizon.

> This is eclipsed by approaches that focus solely on actual experiential content, insofar as it fuels the formation of propositional attitudes, and that construe recognition in a spectatorial fashion, rather than situating it in patterns of affective relatedness. (Ratcliffe, 2008, p.213)

On this account the content "this person is not my spouse but an impostor" (as in Campbell's rationalist account) is part of the experience, and therefore it is not the case that a delusional belief *causes* an experience that is drained of affect, because the affect can actually approximate the content of the delusion. Also, according to Ratcliffe, experience is much richer than just perceptual content because the former incorporates actuality, possibility, practical dispositions and a background sense of reality. Therefore, a second stage in the two-factors is not needed either because the act of reasoning "is something that is embedded in our experience, rather than something to be set apart form it" (Ratcliffe, 2008, p.214).

Finally, Ratcliffe wonders why it is the case that people with Capgras assume the position "entity x is not my wife, but an impostor". His answer is that the experience has such an effect upon their conduct that they cannot override it, regardless of which propositions they utter.

Ratcliffe is surely correct in his observation about the underlying assumptions of the debate about delusions in analytic philosophy. It is true that they share "a commitment to explaining delusions in terms of causal relationships between perceptual and belief contents" (Ratcliffe, 2013, p.229). This reflects the aims of analytical accounts, where the description of the phenomena is thought to be complete when it is judged under those assumptions. The question here is if those assumptions are correct. Just as Jaspers states that the notion of delusion being a belief is not necessarily wrong, one has to wonder if philosophy loses something by not taking into account broader phenomenological traits. It could be the case that analytic philosophical accounts could have a perfectly valid question without focusing on the experience itself. After all, the question whether delusions are beliefs or not has significant repercussions in the way people and clinicians see the phenomenon. And, as Jaspers and Ratcliffe acknowledge, the enterprise of finding phenomenological understanding is complementary to the question whether delusions are beliefs or not.

d) Objections to Ratcliffe

In this section I advance four objections to Ratcliffe's phenomenological model of delusions. The first one acknowledges that Ratcliffe's account is self-consistent. Nevertheless, this makes it too rationalistic for phenomenological purposes. The second

objection is that Ratcliffe cannot provide true understanding of the phenomenon with his notion of "radical empathy". The third objection concerns the fact that the concept of horizon does not shed new light in the nature of delusions, since it is so similar to existing characterizations. The final objection is about the relation between the concepts of experience and belief.

First, is it the case that phenomenological views have arrived at a true understanding of delusions under their own premises? It is clear that Ratcliffe's answer is more thorough and complete than Gallagher's, and perhaps Jaspers'. His account of delusional atmosphere understood as a shift in the horizon of possibilities and understanding through radical empathy offer us a way of undermining the ununderstandability of delusion. Although Ratcliffe does not offer clarification of many of the paradoxes of delusion, for example, why some patients have insight about some instances and then they fall back into their delusion, to clarify their understandability is a valuable project on its own. Radical empathy offers a way to answer the question of what delusions are in a new way whilst maintaining their bizarreness. To the question "what are delusions?" Ratcliffe offers an answer that could be summarised through the statement "delusions are one-factor *experiences*, which are understandable, not through pure logical reasoning, but through empathy, where empathy means a perception-like exploration of others' experiences through personal interaction. The question whether they are beliefs or not is beside the point". This definition is well argued under its own premises, and it offers a genuine alternative to the common false belief based definition of delusions.

However, this prompts a corresponding worry. Is Ratcliffe's explanation too rational? Is his account of Capgras really distinct from those offered in analytic philosophy? Ratcliffe says that the experience of strangeness would make someone feel as if another person is not a person, but an impersonal object. This has to do with the horizon of possibilities that the other (in this case, the spouse) offers, but still, is an *explanation* of the perceptual cause of what creates a delusional thought. Is it the case then, that we are substituting for propositional attitudes the concept of "experience"? If this is true, then the explanation is very similar to Maher's (1974) account of delusions as *rational* responses to an anomalous experience, just that in this case, we will not worry about the sub-personal level cause, but would worry about having a strange experience. But is it not the case that having a strange experience and proposing it as the cause of delusion

is, close to a tautology? Because, to the question "What causes delusions, and what is its nature?" the answer would be that the strange experience itself is the delusion, characterised by anomalies of the horizon of possibilities. No doubt Ratcliffe would say that the *cause* of the delusion is not the central matter to phenomenology, what matters is understanding the experience. However, it is not clear what the explanatory gain is in shifting to Ratcliffe's description.

There is the further question of whether Ratcliffe really achieves understanding; in other words, is "radical empathy" enough for a true understanding? This issue turns on whether radical empathy can really be applied, because, even given the fact that we change our horizon of possibilities, Ratcliffe does not offer us an alternative to mere perception in his account. The main problem is that whilst we can make sense of some modest changes in the horizon of possibilities of objects or other people, the kind of changes that seem to be expressed in delusional utterances are inaccessible to us. Labelling our access 'radical empathy' is not enough to show how this is possible. Ratcliffe wants empathy to be perception-like, but then he does not describe the difference between observing with empathy a human being and observing carefully a cup of coffee. Because we want second-person intentionality to be different from pure perception, his answer would probably be that empathy comes naturally to rational human beings when making an observation, but if this is the case, why would so many people believe that delusions are so bizarre and not understandable in the first place? So the answer cannot be that it comes naturally; Ratcliffe needs a further account of radical empathy to be consistent.

In the case of severe depression, Ratcliffe (2014) proposes a definite account of the change in the horizon of possibilities. That is, the shift of the possibilities in depression is reflected by a lack of action. In the case of Capgras the shift of possibilities lies in the strangeness of the content "This person is an impostor". However, Ratcliffe does not offer us the same rationale for people with schizophrenia. He does seem to gesture at it, as if it should be clear that some change in the horizon of possibilities should be enough, and therefore would bring an alternative reality. If so, that does not help us understand it, because the "change in the horizon of possibilities" has the same features as schizophrenia itself.

Finally, for Ratcliffe, we are charging the concept of "belief" with a role it cannot fulfil because we put in it everything that *an experience* should incorporate, such as actuality,

practical predispositions, possibilities, and a sense of reality. Although the term "belief" is itself valuable to doxastic and anti-doxastic accounts, delusions are much more than this according to phenomenology. But, Ratcliffe's approach accepts that a definition of belief could be helpful to both analytic and phenomenological views: it all depends on the boundaries between belief and experience. Ratcliffe's account gains much of its plausibility by loading some of the features of belief into experience itself. His phenomenology does not so much undermine doxasticism itself, it just opens a wider door for the analysis of delusions.

IV. Conclusion

Phenomenological accounts aim for a different sort of description about the nature of delusion when compared to more analytic traditions. Therefore, the underlying assumptions are different. The most salient ones are:

- The concept of belief is not central to the nature of delusions- Phenomenologists do not deny the possibility of delusions involving beliefs, or another propositional attitude. They just disagree on the importance of the latter. Doxasticism is a possible complement to their definition, not the central matter for its classification.
- The importance of subjective experience- Phenomenologists describe the experience of the subject through concepts such as delusional atmosphere, horizon of possibilities and alternative realities where subjective and personal features are present.
- Understandability- Contrary to Jaspers, contemporary phenomenologists argue that delusions are understandable through a more encompassing concept of empathy.
- Empirically adequate explanations- Phenomenologists such as Gallagher and Rattcliffe maintain that their views are complemented by empirical explanations of the features of the phenomenon. However, it is not clear whether they are aiming at scientifically grounded explanations so much as clearer descriptions of delusions.

Finally, although Ratcliffe's account is more successful than Gallagher's in its explanatory status, neither provides a satisfactory account of the nature of delusions. The main

reason for this comes from the vagueness of the explanatory devices such as alternative realities, or radical empathy. The latter concepts do not explain the main problematic features of delusions, such as their pathological nature, whilst being too vague to reflect the complex nature of the concept of delusion. In other words, phenomenologists play down the problems encountered in delusions to give an account that encompasses the subjective nature more than the specific puzzling questions of delusions. Concepts such as empathy and understanding keep the vagueness inherited from Jaspers. Therefore, although such accounts may fit patients' narratives, the pathological and puzzling nature of delusion remains.

CHAPTER V- BERRIOS, CAMPBELL, AND SASS: CONDITIONS OF POSSIBILITY OF PROPOSITIONAL ATTITUDES

I. Introduction

The preceding chapters have focused explicitly on an ontological question: what sort of mental state are delusions? Different accounts have attempted to shed light on delusions by comparing them to, for example, irrational beliefs or acts of imagination. In each case, however, it has been possible to frame some general objections to them. However, those discussions prompt a further important question about the nature of delusions: would an ontological analysis be complete without asking whether the meaning of the utterances characteristic of delusions is a shared public and hence available meaning? Berrios, Campbell and Sass are concerned with this question and they propose that an analysis of delusions can be approached via addressing their meaning. Berrios', Sass' and Campbell's accounts differ from previous stances which take the meaning or content for granted while discussing the possible functional or phenomenological roles of delusions. Berrios, Sass and Campbell, by contrast, question this.

German Berrios suggests that delusions look like they might mean something, but they are actually empty speech acts. His view is that the important part of the phenomenon occurs before the delusion is crystallized, the pre-delusional state. Delusions "are not the symbolic expression of anything" (Berrios, 1991, p.12). Thus, Berrios maintains that delusion accounts as "wrong beliefs" is misleading; delusions are simply not beliefs.

John Campbell's main concern about delusion has to do with the underlying meaning in the patient's utterances and to what sort of rationality (or irrationality) does that lead: "The key question is whether the deluded subject can really be said to be holding on to the ordinary meanings of the terms used" (Campbell, 2001, p.95). In order to do so, he is concerned with what are the pre-conditions for meaning to have some sense at all in delusions. To answer this, he considers a Wittgensteinian concept: framework propositions.

Louis Sass, whose work often draws on the tradition of phenomenology, also has an analytic-philosophy side taken from Wittgenstein. He proposes that delusions can have

some sort of Jaspers-like understanding by seeing them as a similar doctrine to solipsism, and through this, his account treads a fine line between empathic understandability and what does not make sense any more.

Sass' and Campbell's accounts are embedded in the task of giving sense to what most people would agree it is hard to make any sort of sense of. Their approaches, although different on the surface, share this broader commitment to sense making. On the other side, Berrios argues that there is no content and no information in the delusion. In my opinion, contrasting these opposite views makes the discussion about meaning more encompassing.

Berrios', Campbell's, and Sass' question concerns whether what we call "meaning" is problematic in delusions given that it is not clear what the patients mean by their utterances.

These three accounts are also similar in the following aspect: all work on the assumption that rationality and meaning are deeply connected. Some philosophers disagree with this. In doxastic stances, such as Bortolotti's, rationality does not need to have a prominent place either, since to her, all beliefs are irrational up to a certain extent. Conversely, in Sass and Campbell one sees the need to go further and to inquire about the relation between rationality, meaning and belief, because they think that an account of the nature of delusions would be incomplete without analysis of their meaning or content.

The accounts in this chapter were chosen because they highlight the fact that the content of the delusion is problematic due to the relation between meaning, belief and rationality. Sass differs in an important matter with respect to Berrios and Campbell. The latter's accounts cannot establish what the content of the delusion actually is, whilst Sass thinks that the content can be grasped by looking to the limits of rationality, meaning and understanding in the philosophical doctrine of solipsism.

Through this chapter I will analyse the conditions of adequacy that constrain the assumptions that Berrios, Sass, and Campbell propose in order to question the sense in delusions. The main possible outcomes that are presented in this chapter are the following: Delusions do not and cannot make sense at all like Berrios maintains; delusions lie on a border line of understandability, as Sass argues; or the sense of delusions is incommensurable with everyday sense, like in a paradigm change, as Campbell seems to suggest. Ultimately I argue that the fundamental question is whether

solipsism or deviant framework propositions can make sense, that is, can have meaning, in the first place; and that the notion of an empty speech act does not explain the empirical data.

II. Delusions as empty speech acts: Berrios

Amongst psychiatric clinicians the view that at least some delusions are utterances without meaning is widespread. This view seems intuitive in some manifestations of delusion: for example, in cases of "perseveration[15]" where the patient repeats the same phrase over and over again independently of the context; or when the patient says something that is remarkably bizarre (for example," my wife is a robot"), or plainly impossible as in Cotard delusion (for example, "I am dead and rotting inside"). Furthermore, a psychiatric symptom of the schizophrenia spectrum is the speech of the patient being disorganized (DSM 5, p.88). As in many psychiatric symptoms, the severity may vary, but the acuteness of speech disorganization could be an indicator of the seriousness of the mental disorder when other possible causes are discarded (factors such as a possible stroke, Alzheimer, dementia, Tourette's, etc., have to be ruled out first).

German E. Berrios, a psychiatrist who has written and analysed historical conceptions of mental disorders (Berrios, 1996), is one of the earliest proponents of an account about the nature of delusions that is still debated as a contemporary view, for example in (Thornton, 2007), (Bortolotti, 2010), (Gipps, 2012). His view is that delusions are empty speech acts whose content is thus unimportant because it is lacking. Such an assertion has been qualified as radical but influential from the aforementioned philosophers, and therefore, it is worth analysing its main claims.

As a previous note to the discussion of Berrios' account I will offer my reasons for placing him in this chapter. One of the crucial features of this chapter is analysing how preconditions for the possibility of belief are problematized. Berrios' contribution meets this criterion, which is shared with the other two authors analysed: Louis Sass and John Campbell.

[15] Perseveration can be defined as the contextually inappropriate and unintentional repetition of a response or behavioural unit. In other words, the observed repetitiveness does not meet the demands of the situation, is not the product of deliberation, and may even unfold despite counterintention (Crider, 1997, p.63).

Berrios' positive claim problematizes the assumption that delusions are content-bearing states. His claim is that delusions are empty speech acts, and the main argument he gives for this claim is that they are not beliefs. The argument he presents for the latter (delusions not being beliefs) is secondary. It is in view of his positive claim, rather than the argument that he presents for it, that he most appropriately belongs to this chapter. Berrios' rejection of doxasticism would be a valid argument for his positive account if the only way to issue a genuine speech act (one with content) is by expressing a belief. This would eliminate all the alternatives to the empty-speech-act hypothesis. But of course we express all kinds of attitudes, besides belief, in genuine speech acts. Berrios' argument doesn't eliminate these other alternatives, and so is not a valid argument as it stands. Therefore, Berrios' account is best discussed in this chapter.

a) Berrios' negative account: delusions are not beliefs

Before presenting Berrios' account of delusion as empty speech acts it is worth analysing his reasons for rejecting the view that delusions are beliefs. Instead of offering reasons of why an empty speech act sheds light on the nature of delusions, he takes the burden of the argument to be carried by his dismissal of the belief stance; given that Berrios considers other historical alternatives erroneous he does not feel the need to illustrate why he thinks that delusions are empty speech acts. Therefore, I think it is useful to analyse his antidoxastic arguments first.

Berrios, in his (1991) paper, shows how the concept of delusion had a parallel philosophical development along with the notions of insanity and belief. Some views, from as early as the seventeenth century, proposed that to have delusions was to be insane, and vice versa. However, modern concepts of belief began only in the nineteenth century. Then, the concept of delusions as wrong beliefs began when both concepts acquired its modern character: "the view that delusion is a subtype of belief started during the nineteenth century and was a consequence of the conceptual separation between knowledge and belief" (Berrios, 1991, p.10). It was during this historical moment that the choice about the nature of delusions had to be discussed in those terms: it was clear that delusions did not fit very well the definition of belief; they were "wrong beliefs"; however, a belief could be "wrong" in two different ways, it could be

wrong about its form or about its content. Berrios explains what the different implications are:

1) The form of the belief is wrong- Delusions are "misshapen structures" (Berrios, 1991, p.8). This means that they are anomalous and do not conform to any structure. I take that Berrios' position is closer to delusions being sui generis rather than non-paradigmatic beliefs, as in Reimer[16] (2010), in this case.
2) The content of the belief is wrong- The content of the utterance is false in the sense that is contrary to reality.

Berrios implies that it was at this historical point where the choice about the status of delusions was mistaken and led to modern misconceptions. A particular example is Jaspers:

> Confronted with this choice, Jaspers opted for the latter: he was convinced that delusions were "structurally sound" beliefs whose "content" was discrepant with reality. The latter was due to their pathological (morbid) origin. By reasserting this 19th century view, he effectively led the descriptive psychopathology of delusions down a blind alley. (Berrios, 1991, p.8)

What Berrios suggests with this discussion is that other options, different from the doxasticism of delusions, were available in the past but they did not receive the necessary attention. Berrios then proceeds to test whether delusions are beliefs with a model taken from Price (1934). In this model, the elements that comprise a belief (P) are (Berrios, 1991, p.8):

a) Entertaining P, together with one or more alternative propositions Q and R.
b) Knowing a fact or set of facts (F), which is relevant to P, Q or R.
c) Knowing that F makes P more likely than Q or R, i.e. having more evidence for P than for Q or R, and
d) Assenting to P; which in turn includes i) the preferring of P to Q and R, ii) the feeling a certain degree of confidence with regard to P.

Berrios' argumentation that delusions do not fit a), b), and c) is brief. He only states that those criteria are regularly absent in clinical observation. But he does not cite any

[16] I discuss Reimer's account in chapter III of this thesis as a doxastic view.

particular observations or cases, nor elaborate his reasons any further. About d) he argues that a subject that truly believes P does it according to certain probability constraints which could be related to evidence, personality, emotions, etc., which patients do not show. Berrios suggests that the assumption that delusions are beliefs led to accounts, such as Hemsley and Garety's (1986), that propose that Bayesian probability might shed light on the origin of delusions, but that these studies are not helpful because:

> Apart from the general point that an assumption of rationality (i.e. the mind of a statistician) is not warranted in the case of psychotic subjects, and of the specific point made above that delusions may not be beliefs at all (…) there is some evidence that, after all, "normal" subjects do not often observe Bayesian inferential rules either. (Berrios, 1996, p.114)

Berrios concludes that delusions are not like normal beliefs, and then "it must be asked, why persist in calling them beliefs at all?" (Berrios, 1996, p.114). Berrios, then, presents his theory of what delusions are, which I discuss in the next section.

b) Berrios' null hypothesis: delusions as empty speech acts

i. Berrios' historical rationale

The best source to understand Berrios' account of delusions is his (1996) book, *The History of Mental Symptoms*, where he presents a comprehensive compendium of the main accounts of delusion that go from the seventeenth to the twentieth century. His chapter devoted to delusions serves two purposes: first, it is an interesting analysis of the development of the concept, but more importantly, it presents the basis on which Berrios offers his assertion that delusions are empty speech acts.

It is very early on in the chapter that Berrios makes his position clear, "properly described, delusions are empty speech acts that disguise themselves as beliefs" (Berrios, 1996, p.114). Given the bizarreness of delusions and the anomalous behaviour that patients sometimes present, Berrios' assertion is not entirely surprising; nowadays

many clinicians hold this view. It seems to come naturally to people in direct contact with patients. However, Berrios' reasons for maintaining his positive view come from the early nineteenth century.

Berrios observes that, by the 1850s, complex classifications of insanity emerged. This was promoted by the psychodynamic movement, which took away the attention from an important phenomenon: the "pre-delusional state". Pre-delusional state is, "the psychopathological events immediately *preceding the crystallization of the delusion*"(Berrios, 1996, p.115). Berrios suggests that delusions are just the culmination of a severe mental pathology that has been going on in the background for a long time. Berrios admits that the pre-delusional state is subtle and hard to identify because it does not have strong defining symptoms as paradigmatic accounts of delusions do. He describes pre-delusional state as an all-pervading, ineffable pathology that prevents the patient from reaching reasonable conclusions. However, Berrios claims that, even if this state is ineffable, it does not mean that it cannot be conceptualized. He suggests that pre-delusional state is an expression of distress in a specific brain module, and that this is the actual mechanism of delusion generation.

It is not quite clear to Berrios which of the contemporary models would be adequate for explaining pre-delusional state. Some of the candidates that he considers for the latter task are Maher (1974) with his view of delusions as "anomalous experiences", Schneider (1959), who states that delusional mood precedes delusional perception but it does not cause the delusion, and Conrad (1958) who highlights the ineffability of the phenomenon, amongst others.

Berrios, credits Austin (1962) with the term that delusions are meaning "masquerades". He cites many other historical accounts of delusion that suggest that delusional mood or similar concepts could complement the study of the nature of delusion. Some of the examples described by Berrios are Cotard (1880), Bleuler (1906), De Clerambault (1909), Hesnard (1924), and Marco-Merenciano (1942), amongst others.

It is interesting to note that the concept of delusional mood emerged before Jaspers, in Hagen (1870). Hagen suggests that the symptoms of delusional mood are usually eclipsed by the severity of the delusion and the fact that clinicians do not usually investigate it. Berrios quotes Jaspers without giving any further explanation of what he thinks of his description of delusional atmosphere. I reproduce here a part of that quote:

> If we try to get some closer understanding of these primary experiences of delusion, we soon find we cannot really appreciate these quite alien modes of experience. They remain largely incomprehensible, unreal and beyond our understanding (…) the use of the word 'atmosphere' might suggest psychasthenic moods and feelings perhaps and be a source of confusion; but with this delusional atmosphere we always find an 'objective something' there, even though quite vague, a something which lays the seed of objective validity and meaning. (Jaspers, 1913, p. 98)

Although Berrios does not say it explicitly, in my opinion, pre-delusional state is similar to Jaspers' delusional atmosphere[17] in the sense that a) it is a compulsory stage before the delusion, but not causal, and b) it contains the first subtle pathological symptoms. However, although in the first part of the quote above Jaspers affirms that delusions are alien modes of experience and beyond our understanding, what Jaspers means is quite different from Berrios' account. The first difference is that Jaspers' delusional atmosphere is not the most important part of the delusion, as it is for Berrios: in other words, for Jaspers, primary delusions (not delusional atmosphere) are the pathological part of delusion, and to Berrios, the information about the patients' state of mind is in the pre-delusional state. The second difference is that for Jaspers the delusion is an expression of what the patient is experiencing, and it has meaning in the sense that it reflects what the patient is going through, and although one cannot empathize with the delusion (it is ununderstandable), it still gives information about the phenomenological experience; to Berrios delusion per se gives no information at all.

Berrios observes that during the early twentieth century delusions were classified according to their content: nihilistic, persecutory, erotomanic, grandiose, etc. Also, classifications continued with specific states like Capgras, Fregoli, and others. He seems to imply that these classifications are not useful when he affirms that "some of which, thankfully, have by now been forgotten" (Berrios, 1996, p.125). In contrast, Berrios finds that nineteenth century reports are more accurate than the above mentioned classification because in the nineteenth century delusions were described with features like imperviousness to external pressure, insight, dissolution and conviction, amongst

[17] Jaspers' delusional mood is discussed in Chapter I of this thesis

others. These characteristics, according to Berrios, might bear clinical and neurological meaning. This idea, combined with Berrios' claim that the important part of the phenomenon is the pre-delusional state result in his affirmation that the content is not relevant for making sense of the nature of delusions and therefore, delusions are empty speech acts.

ii. Empty speech acts

Berrios argues that the information that delusion conveys comes *before* the delusion is crystallized. When the delusion is already formed, what is left is only a random piece of information.

> The content of delusions is inscribed in a different discourse and reflects cultural and personal codes: a consequence of this claim is that crystallized delusions become an interesting detritus, a 'foreign body' the explanation of which is not necessarily linked to that of delusion formation. (Berrios, 1996, pp.124-125)

Therefore, Berrios' assertion that delusions are empty speech acts is related to his arguments concerning delusional atmosphere; if the real pathology occurs before it takes the form of a specific delusion, the delusion itself and its content are not important. Delusion is just an accident in which too many factors are involved so that any useful neurological information is already lost.

> [D]elusions share no structural similarity with normal beliefs (…) they are best described as 'empty speech acts'. Delusions are called 'empty' not because they have no content (they all do!) but because they have been 'shelled out' of all information in regard to the brain address where they were formed, and because they no longer can be integrated or 'inscribed' in the pragmatics of a discourse between doctor and patient. (Berrios, 1996, p.124)

Berrios also observed that the rejection of the "all-or-none" model[18] of delusions "has led to the unwarranted claim that a "continuum" might then exist between overvalued ideas, obsessions and delusions" (Berrios, 1996, p.125-126). Berrios' latter comment is challenged by Bortolotti (2010); I present her arguments in the criticisms section of this chapter. Finally, Berrios finishes with the remark that delusions do not constitute a suitable explanation for any interesting matter:

> Furthermore, the bizarre and contradictory clinical nature of some delusions, would make it unlikely that they are actually explanations for anything, or indeed that they are fulfilling an adaptive or teleological function. (Berrios, 1996, p.118)

Although Berrios' account is radical, it is still widely debated in philosophy of delusions due to the fact that it is a possible option that seems intuitive, and that many clinicians are inclined to it.

c) Criticisms of Berrios' account

Berrios' suggestion that delusions are empty speech acts is backed up by the claim that delusions are not beliefs, as I describe in the previous sections. Criticisms have been raised both against the former view and against the antidoxastic argument he gives in defence of it. I will start out by looking at criticisms of Berrios' argument that delusions are not beliefs, after which I shall turn to look at criticisms of Berrios' claim of delusions as empty speech acts.

i. Bortolotti- against antidoxasticism in delusions

Bortolotti's arguments are mainly against the model of belief that Berrios endorsed in order to dismiss the doxastic status of delusions. This sort of criticism is important in order to counter Berrios' base for dismissing contemporary accounts of delusion, as well as analysing what Berrios thinks as what carries the burden of proof of his view. In

[18] The "all or none" model in delusions is present in accounts in which patients are considered to have or not a certain skill or symptom regarding diagnosis. The characteristic in question cannot be said to be partially held. Particularly, insight is said to be completely absent in delusions in these models (Cfr. Gibbs and David, 2003).

Bortolotti's view, Price's criterion[19] is too narrow to define belief, but since Berrios' account has proven very influential, she gives special attention to his remarks.

First, Bortolotti notices that even if Price's conditions are not met, the possibility of it being a different intentional state, such as an imagining, could still be the case. These other intentional states do not need to be attached to any sort of probability in the first place.

Second, Bortolotti poses a problem for the assertion that endorsing a statement which has less probability than others makes this statement a non-belief. In order to make her point, she suggests the following model about probabilities:

> Here is an example of the standard process of belief formation according to Berrios. I went shopping and got some cookies. When I get home, I find no cookies in my bag. Why are there no cookies? Different potential explanations come to mind. I intended to get cookies, but then I forgot. Another possibility is that I got the cookies and paid for them at the cashier, but forgot to pack them with the rest of the shopping. A third possibility is that my evil neighbour made the cookies disappear with a spell. The fact that there are no cookies in my bag has many potential explanations which can be more or less plausible in context. Berrios assumes that if I seem to endorse an explanation that I do not regard as the most probable, then I cannot genuinely *believe* that explanation. (Bortolotti, 2010, p.119)

In delusions, the subject might endorse an explanation that is not the most probable one. Then, according to Berrios' view, the subject does not believe what he endorsed. Then, if deluded subjects do not believe in the delusion's content, "why do they behave as if they believed something they do not believe? Are they deceiving others or themselves?" (Bortolotti, 2010, p.120).

Third, about the assertion that deluded subjects do not look for evidential support, Bortolotti argues that there are examples from Capgras cases in which subjects maintain the hypothesis that their spouse has been substituted as an explanation for an abnormality observed in the spouse. So, the subject might have a reason to claim that

[19] Price's (1934) criteria are quoted in a previous section of this chapter

his spouse has been replaced; it might not be a good reason, but it is different from having no reason at all; in other words, it might not be a sound reason, but it is not completely ungrounded.

Fourth, Bortolotti argues that delusions are not *rational* beliefs, but subjects with normal beliefs do not necessarily choose the most probable belief either. Religious beliefs are an example in which this seems to be the case. I would add to Bortolotti's point that subjects that gamble are a clear example of this instance too.

Berrios could respond to Bortolotti that delusions are not a different propositional attitude because they are empty speech acts, and these are more like "cognitive noise" rather than some other intentional state. Bortolotti's counterargument would be that cognitive noise would not affect behaviour, be endorsed, and purport emotional content as delusions sometimes do.

To counter Bortolotti's second argument, Berrios could say that, although it is true that subjects sometimes behave according to the delusion, there are also instances in which they do not, and that those are the cases where delusions are empty speech acts. Bortolotti in this case could say that normal beliefs do not necessarily go with the expected behaviour, as in cases of common hypocrisy, or failures in predicting self-performance.

About Bortolotti's third argument, Berrios could argue that rationality can only be achieved with good reasons, however in this case, it would be a matter of what one is willing to consider rational. What happens here is that Berrios' criteria would leave many common mistakes out, but Bortolotti's would allow delusions *and* normal beliefs to be irrational. In my opinion, the matter of what counts as rational would have to be solved first. About the rationality of normal beliefs, Berrios explicitly concedes Bortolotti's fourth point: "there is some evidence that, after all, "normal" subjects do not often observe Bayesian inferential rules either" (Berrios, 1996, p.114).

In general Berrios' argument is not sufficient to deny delusions a belief status. In the following section I consider criticisms against his positive claim.

ii. Objections against the lack of meaning in delusions

The main problem with thinking that delusions are empty speech acts is that this account does not do enough to say how delusions sometimes connect to other utterances and other actions. It is true that at times delusions do not engage in the corresponding behaviour, and it might look as if Berrios' account is justified in those cases. Therefore, I will consider both cases: a) when delusion is connected to speech and behaviour, and b) when it is not.

In the case a) when delusion is connected to speech and behaviour, the view of it being an empty speech act is clearly unjustified. It is mainly through *the patient's behaviour and speech related to his delusion* that it can be seen that something is wrong with him. People with paranoid schizophrenia who say that the CIA is after them act with distrust towards their own friends and family, and when asked about it, they will assert that their actions are being monitored. If the delusion were just an empty utterance then how could his actions be explained? There are cases in which patients have done very extreme things guided by their delusion. In some Cotard cases the patient affirms that he is dead and so he wants to be buried (Debruyne, 2009). In the Capgras case that I quote at the introduction of this thesis, the patient says that his wife has been replaced by a stranger who looks very much like her, but it must be noted that the patient says he is worried about the safety of his "real" wife and acts towards the alleged stranger (actually, his wife) as if he did not know her. These examples show that not only patients act and endorse their claim; in many cases there is a net of beliefs attached to the delusion that most of the times result in some sort of action.

In the case b), when the patient's behaviour and speech seem disconnected from the world, Berrios' reasons are unhelpful. Berrios says that rationality (specifically demonstrated by being competent in statistics) cannot be assumed in deluded patients. It has been noted that deluded patients are equally or more competent than normal subjects (Huq et al, 1988). Regarding inferential reasoning, researchers have found that "differences in reasoning between deluded patients and controls are surprisingly small" (Kemp et al, 1997, p.398), therefore, Berrios cannot claim that patients are irrational in this way either.

It is interesting that Berrios is focusing on what the meaning of delusions could be because it is a puzzling aspect of delusion that looks outside its characterization and

aetiology, the latter being the most discussed aspects of delusion. So, discussion about meaning in delusions is important and welcome. However, Berrios' conclusion is more like a null hypothesis: that they are empty of meaning and that they are to be modelled on a speech act that completely fails. This naturally implies that delusions do not have any rational connections and, therefore, there is nothing to be made sense of. Still, Berrios does not attempt to explain what he understands empty speech acts are, and neither in which way delusions belong to this sort of state. Berrios' main drive behind his view seems to be that he recognized an available category which seemed promising and is intuitive to clinicians. When he proposed his view of delusions as empty speech acts, Berrios also seemed to have in mind that delusions are not informative regarding aetiology. I quote his assertion again:

> Delusions are called 'empty' not because they have no content (they all do!) but because they have been 'shelled out' of all information in regard to the brain address where they were formed. (Berrios, 1996, p.124)

However, that delusions do not give information about the brain does not mean they do not matter. Thornton (2007, p.102-103) observes that Berrios' claim comes from a particular philosophical view that explains the content of mental states as a "bottom-up" view, where in this specific case, being bottom-up implies that all the information comes from brain "modules" where meaning is atomic.

I would argue that, in a very simplistic way and given the above view, it can be the case that a faulty brain module causes something as an utterance that the patient does not mean. However, putting more pressure on this view, it is clear that delusions do not work that way. The fact that patients have other non-faulty beliefs shows it is not as simple as delusions being caused by a faulty belief module. With respect to Berrios' characterization of delusions as "a random fragment of information "trapped" in the very moment the delusion becomes crystallized" (Berrios, 1991, p. 12), it could be argued that a random fragment of information, in any case, could be easily dismissed given that it would not have the support from other thoughts that intervene whilst learning something about the world. That people have random fragments of information is not uncommon; sometimes normal people have irrational thoughts or make mistakes,

but these do not persist like delusions do, precisely because they are not backed up by other beliefs.

A more fundamental objection to Berrios' account is that speech is capable of expressing a whole range of things, not just beliefs. When Berrios dismisses Price's account of belief and then implies that delusions are empty speech acts this is a non sequitur because he then has overlooked other accounts that show there are many propositional attitudes available. In other words, as I affirmed earlier, Berrios' dismissal of doxasticism would be a valid argument for an empty speech act theory only if the only way to issue a genuine speech act would be by expressing a belief. Therefore, it is clear that there is a gap in Berrios' argument in this sense given all the accounts I have analysed previously in this thesis.

An objection against Berrios' claim that content is unimportant is that some schizophrenic patients try to befriend the voices they hear in their heads, and this has proven useful to some extent because it helps patients to reduce stress. This indicates these patients engage with the delusion at the level of meaning, and sometimes, they take agency of their situation too.

Also, people with delusions are aware that their delusion might sound odd to others:

> [I]f you ask "what would you think if I told you my wife had been replaced by an impostor" [to a delusional patient] you will often get answers to the effect that it would be unbelievable, absurd, an indication that you had gone mad. Yet these patients will claim that, none the less, that is exactly what has happened to their own relative. (Young, 1998, p.37)

So, many delusional patients show awareness that their claim is bizarre to others, and they usually maintain that what is happening to them is unique.

As a final comment, the question about the patients' treatment remains. How should one deal with an empty utterance? Berrios' answer might be that the useful information is to be found in the pre-delusional state, but then, how could information be obtained from the patient if such a state is ineffable and no clear symptoms can be observed?

Berrios is an example of someone who takes thinking about the meaning of delusions very seriously, although his claim is a radical and implausible view. In the next sections I present two authors with more sophisticated views about meaning, who nevertheless

share with Berrios the idea that ascribing meaning to delusions is problematic, although they disagree with Berrios' null hypothesis that in delusions there is nothing to be made sense of.

III. Campbell's approach to delusions

a) Campbell's connection of meaning and rationality

In the first section of his (2001) paper, Campbell is primarily concerned with the question of the relation between rationality, meaning and the content of the delusion. His view is compatible with the Davidsonian constitutive principle of rationality. It is in accord with this principle that Campbell tries to make a diagnosis of the possible sense or meaning in delusions. In the second part of his paper Campbell deploys his positive account of delusions as Wittgensteinian framework propositions. This will be analysed in the next section of this chapter, where I will argue that there is a tension between that account and his assertions about the link between meaning and rationality.

In order to say what the conditions for meaning of a delusion comprise, Campbell develops an account of monothematic delusions that he considers to be in opposition to Ellis and Young's (1999) model. He labels his own approach "rationalistic" and Ellis and Young's approach "empiricist". Therefore, Campbell begins by explaining both philosophical positions:

> On what I will call a rationalist approach to delusion, delusion is a matter of top-down disturbance in some fundamental beliefs of the subject, which may consequently affect experiences and actions. On an empiricist approach, in contrast, delusion is a rational response to highly unusual experiences that the subject has, perhaps as a result of organic damage. Ellis and Young (1990) recently provided an empiricist analysis of the Capgras and Cotard delusions. (Campbell, 2001, p.89)

Campbell begins his argument by stressing the importance of rationality to the relation between belief and meaning. In the case of Capgras, one has to begin by ascribing *the correct meaning* to the characteristic assertion that the allegedly familiar person is,

actually an impostor. Although the person says “That woman is not my wife”, according to Campbell, as I will explain below, a better suggestion for what the person really means is “that [remembered] woman is not my wife”, therefore, “That [currently perceived woman] is not that [remembered woman]” (Cfr. Campbell, 2001, p.90).

Campbell’s approach is constrained by the underlying assumption that rationality and holism must be present in an explanation of the patient’s delusion, and this leads him to question the assumption of a shared meaning between the patient and the public in general. The latter is displayed by Campbell’s particular approach to the question of meaning, which suggests that the failure of rationality made by the deluded individual *is different in kind* to normal, everyday mistakes. For example, there is a sense in which one understands when someone makes a mistake about the capital of a certain country, but it is different to the mistake that a person with Capgras makes. For Campbell, the latter cannot be accommodated in a rational framework. This is because one cannot say what the content of the expression of delusion is in the first place.

Campbell tries to outline some possible options that might render the meaning of the content of the delusion understandable. His first approach is with the standard quote from Capgras patients “This person is not my wife”. A way in which this could be understandable might be that what the patients actually mean is “I am not legally married to that person”. But this is not what the Capgras patients are trying to communicate; they would deny such interpretation because they seem to be worried about their spouse being an impostor, not about the legal status of their relationship. A closer approximation to the actual meaning could be something like “This perceived woman is not that remembered woman”. However, this fails too because, in order to *mean* that, patients would have to engage in the usual paradigmatic ways of finding out if this is true, and there are plenty of cases where this is not the case. So, for Campbell, what patients *mean* is not commensurable; the meaning has to change in order to preserve the relation between it, belief and rationality. However, the price one has to pay to preserve this sort of partial rationality is that the latter conclusion leaves the actual content of the delusion unresolved. So, can the link between meaning and rationality be preserved in this case? True success in saving rationality would involve articulating what is really meant. To say that we, as external agents do not understand some determinate content that the patient wants to communicate is a failure of the Davidsonian framework that Campbell himself is deploying.

Philosophers, such as Berrios, are sometimes tempted to say that this kind of assertion is just an "empty speech" act, but Campbell maintains that the problem with such an account is that patients are making sincere assertions, and they seem to understand what they are saying. Therefore, this suggests that we have something to explain. Two possible options, amongst others, are:

1) Empiricist- The patient is rational and understands the meaning of the words that are being used. The delusion is just a bottom-up, rational reaction to an unusual experience (Ellis and Young, 1999).
2) Rationalist- There is a top-down disturbance in the patient's belief, and the question whether the patient understands the meaning of his statements has to be explained (Campbell, 2001).

Campbell presents his objections to the empiricist view first in order to make room for his rationalist account. As Ellis and Young present their proposal, the idea is that the patient attempts to explain her anomalous experience by supposing that the person of whom she has memories has been replaced by an impostor.

But it is actually very hard to see how this could be a rational reaction to the patient's experience. Indeed, the experience does call out for explanation, but that hypothesis could not be sustained. For that hypothesis has to be checked by asking whether the observed person is indeed the one of which you have memories; all of these further enquiries show that this is not in fact an impostor.

Campbell ends his attack on empiricism with the question of how a particular experience may lead to the radical implications to which delusions give rise. He explains his philosophical query with the following reflection:

> A simple way to bring this out is to reflect on one of the classic examples of delusion: a patient who looked at a row of empty marble tables in a café and became convinced that the world was coming to an end. (…) The problem is to understand how any experience at all, let alone an experience of marble tables, could be relevant to the verification of the proposition, "The world is ending". What we cannot find is any relation between the patient's use of the words and his knowing of their references. (Campbell, 2001, p.95)

The content of the delusion, Campbell, says, has to be a matter of a top-down inference, because the meaning of the terms used are not sensory terms: "It is not as if the patient has been taken in by a simple illusion of colour or motion, for example" (Campbell, 2001, p.96).

b) Campbell's framework propositions approach

In this section, I will be concerned with the second theme of Campbell's arguments. In his positive account, Campbell claims that monothematic delusions can be seen as what Wittgenstein calls "framework propositions" or "hinge propositions": "The delusions of the Capgras or Cotard patients may be relatively circumscribed in their significance, but within the appropriate domain of enquiry, they may, nonetheless, be functioning as framework propositions" (Campbell, 2001, p.97) In his book *On Certainty* (1969), Wittgenstein explains that propositions such as "This is one hand and this is another", "The world has existed for quite a long time", and "There are a lot of objects in the world" are not ordinary factual beliefs. They are more like the background on which other propositions can make sense:

> All testing, all confirmation and disconfirmation of a hypothesis takes place already within a system. And this system is not a more or less arbitrary and doubtful point of departure for all our arguments; no, it belongs to the essence of what we call an argument. The system is not so much the point of departure, as the element in which arguments have their life. (Wittgenstein, 1969, p.105)

Campbell thinks that a parallel can be drawn between hinge propositions and delusional thoughts. Therefore, propositions such as "My neighbour has been replaced by an impostor" or "I am dead" would have the same kind of status as "This is one hand and this is another". Campbell observes that such propositions are not subject to empirical scrutiny. Hinge propositions can also be about specific matters such as "there are tables and chairs in this room".

Rationalists' views have the burden of proof about the anomalous experiences associated with delusions through an a priori thought. Campbell maintains that the lack of affect observed in Capgras, and the depression observed in Cotard patients, can be

explained by stating that the hinge proposition is generating those affective states. Paraphrasing Campbell: If you genuinely were to believe that your spouse has been replaced, you would of course feel a different affection towards your spouse: the hinge proposition explains the change in the affect. In Young, it is the different affect that generates the strange thought (Cfr. Campbell, 2001, p.97). Campbell makes this move in order to preserve the patient's rationality by trying to explain the possible structural form of the delusional belief. However, the fact that Campbell maintains that delusions have a different meaning for the subject from what is actually said in a shared language puts his own Davidsonian view under tension. He cannot state in shared language what the content of the deviant framework position is.

Campbell sees the problem of ascribing meaning to the expression of the Capgras delusion and offers a solution: to make sense of them one has to consider delusions as "framework propositions" in a Wittgensteinian way. If this were the case, the meaning would have to be taken for granted, and many future actions and spoken justifications would follow from there. What the patient means would still be unresolved, but one could "bracket" the meaning and one could just focus on the consequences of such an assertion. This suggests, as I mentioned earlier, that Campbell's account cannot be classified as a straight-forward doxastic account. What sort of belief is it that has an indefinable content? What sort of belief is a deviant hinge proposition, anyway?

In order to explain later why it is so strange that one could have a deviant framework proposition in the way Campbell maintains I will analyse briefly what a framework proposition's Wittgensteinian role is.

Although framework propositions look epistemological, that is, they look as if they are making statements of the conditions of knowledge, their true role is to give rules that help to define the meaning of words. I will clarify this with McGinn's (1989) account of the matter. McGinn observes that framework propositions, which she calls "Moore-type" propositions, are not homogeneous regarding their content, since they can be about any topic; they constitute some sort of motley. The fact that they are not objects of doubt or knowledge serves to reveal that they are of a special kind. Their role is to constitute a system of judgements which a community accepts without question; a final end to our knowledge-questioning and assertive claims. She highlights their non-epistemic status:

> None of this can be used as a basis for establishing knowledge claims involving Moore-type propositions, and Wittgenstein does not attempt to use these points in this way. Rather, these points are used as part of an overall argument that reveals the error of the sceptic's initial- and fatal- supposition that our relationship to Moore-type propositions is to be construed as an epistemic one. (McGinn, 1989, p.138)

The fact that these assertions cannot be doubted shows that their essence belongs to language games instead of the empirical realm. Hence, they can become a criterion of linguistic mastery because they play the role of fixing our empirical concepts.

Having shown these considerations, it is unclear what sort of role Campbell argues for delusions as framework (Moore-type) propositions. In order to explain the counterarguments against Campbell I will divide them by theme into two sections. The first section contains arguments against the possibility of meaning in Campbell. The second section will show counterarguments against his positive account of delusions as framework propositions.

c) Objections against Campbell's connection of meaning and rationality

Bayne and Pacherie are also concerned with the question of meaning in delusions. Although it seems that their main concern is to refute Campbell's rationalist view, they do this via attacking his assertions of meaning. Therefore, their arguments are worth analysing here.

Bayne and Pacherie's (2004) first objection against Campbell's account is centred on his claim that the mere lack of affect could not cause the perception's particular content. As a response to that claim, Bayne and Pacherie think that the lack of affect can generate the experience of the lack of affect in the Capgras case. The experience generated is not just that the normal affect is missing; what is actually generated in the patient's mind is a sense of unfamiliarity and estrangement.

In order to refute this top-down argument, Bayne and Pacherie observe that there are two issues to consider first:

a) Is it necessary to accept Campbell's characterisation of the patient's perceptual state?

b) Is it true that a perceptual state could not acquire its content other than in a top-down way?

Before starting to discuss whether Bayne and Pacherie's arguments are successful against Campbell, I would like to highlight that Campbell's description of the aetiology of the perceptual state in patients is unclear, except for the fact that it has to be generated via their thoughts in the first place rather than the other way around. How this actually works as a mechanism is not explained in any detail in his (2001) paper. (His rather functionalist explanation of delusion and brain activity is better deployed in Campbell's (1999) paper, but I will not analyse it here because its main philosophical question is not clearly related to the main theme of this chapter.)

In order to explain Bayne and Pacherie's argument against a) I will first differentiate two of their key concepts: Explanationism and endorsement concerning perception. Explanationism is the view in which perception is given sense (explained) via some desired virtues that are meant to ground it: In this model, the patient adopts the delusional belief in an attempt to explain why he has this unusual experience whenever he looks at his spouse (Bayne and Pacherie, 2004, p.3).

Regarding Bayne and Pacherie's view about endorsement concerning perception, this position is basically some sort of theory ladenness about perception; that is, that perception is preceded and modified by underlying relevant thoughts. Bayne and Pacherie explain what it means to them:

> [W]e prefer the endorsement version of the two-factor approach, according to which the patient's visual perception has the content "This [perceived] woman is not the person who I think of as my wife," or something very close to this. (Bayne and Pacherie, 2004, p.4)

Therefore, in their argument a), concerning the characterisation of the perceptual state, Bayne and Pacherie maintain that two-factor theorists of delusion do not necessarily have to agree with Campbell. First, they observe that such theorists would have the explanationist and endorsement options available. An explanationist version of delusions would claim that the content of the patient's experience cannot be as rich as Campbell maintains, that is, "That [currently perceived woman] is not that [remembered woman]". So, to Bayne and Pacherie, the explanationist version would be something like

"This person looks a bit strange". Alternatively, the content of the experience in an endorsement version would be something similar to "This [perceived] woman is not the person who I think of as my wife". Both are unlike the characterization that Campbell maintains. Therefore, one does not necessarily have to accept his characterisation, since there are other two options available, at least. However, this cannot be the end of the argument because by making this move, Campbell loses the burden of proof because in these two new characterizations it still has to be explained how the particular perceptual content could be developed without inheriting it from the (top-down) belief. Therefore, option b) has to be analysed.

To answer b), Bayne and Pacherie deny Campbell's suspicion that the content of the delusion could *not* be "This [perceived] woman is not that [remembered] woman". Their argument goes like this: Campbell believes that we tend to think about our loved ones in terms of our episodic memories of them, not as the persons that cause a certain emotion in oneself. Bayne and Pacherie affirm that Campbell's judgement is too hasty, because it has been shown that the Capgras delusion only affects close relatives considered as loved ones. In any case, episodic memories might be important, but only because they work as cues for our affective memories. And therefore, this sort of content is not necessarily top-down, as Campbell suggests. Bayne and Pacherie assert that memories of relatives are typically affect-laden, and so, one could actually have the content "This woman is not my wife", where the term "wife" is cashed as "the person that produces that affective response in me". Campbell might respond that a reference to a particular individual might be too specific to enter on a perceptual claim. Bayne and Pacherie argue that the latter does not need to be the case with the following example: suppose that a pair of identical twins, called Jules and Jim, exist. One can tell Jules and Jim apart just from looking at them. However, the way one can do this might not be available from introspection. Therefore, Bayne and Pacherie conclude: it is possible to have a perceptual state warranted by a non-introspectable mechanism in which a particular individual is singled out.

Bayne and Pacherie's second objection against Campbell is that he argues that empiricism cannot explain how the deluded patient arrives at a bizarre conclusion given the fact that he ought to revise his own beliefs; in short, that the patient should verify the content in a rational way given the fact he is considered to have a rational response to a strange experience. To Campbell, the patient "seems to have lost his grip of the

meaning of the word" (Campbell, 2001, pp. 90-91). As I argued in the first section of this chapter, Campbell works under the assumption that belief and meaning are linked by the constitutive principle of rationality, and his main argument against empiricist views is that they fail to reflect such a link in an appropriate way.

However, Bayne and Pacherie think that Campbell's claim which I have just described has four argumentative flaws. These flaws can be summarized in the following assertion: patients tend to follow the constitutive principle of rationality, up until the point their delusion makes that impossible anymore. And Bayne and Pacherie argue that the suspension of meaning would not fit with the empirical data one gets about the patients' claims and behaviour. In other words, Bayne and Pacherie agree with the rationality principle that Campbell defends in the first place, but they do not agree with him in the fact that following the principle necessarily implies that meaning is lost in the case of delusions.

Bayne and Pacherie's argument 1) is that, if the memories of our loved ones are affect-laden, the actual way of verifying whether the person in question is one's wife would be to check if the feeling of unfamiliarity towards that person continues, and this is, presumably, what the patient does according to Bayne and Pacherie. If so, evidence is looked for by the patient, and their rationality is not necessarily damaged, as Campbell claims.

Argument 2) against Campbell is that he fails to distinguish two reasons why Capgras patients might not engage in canonical ways of verification. The first possible reason is that patients lack *inferential competence* (their italics). Bayne and Pacherie disagree with Campbell because patients seem to be capable of drawing correct inferences in other situations.

The second possible reason is that patients are competent in such a way, but that they are not willing to use this skill. This unwillingness could be explained by the fact that the Capgras patients also seem to have paranoid or persecutory tendencies, and they see no case for talking to an impostor who is already trying to deceive them. This unwillingness could also be explained if the patient has a motivational deficit, instead of an inferential one.

Argument 3) concerns the fact that patients show that they mean their utterances literally by acting according to their delusional statements (Cfr. Young, 1999). This can be illustrated by the fact that, in some cases, patients act violently towards the impostor:

in 18% of the 260 cases of Capgras analysed, the deluded acted violently towards the alleged impostor (Forstl et al., 1991). In one particular case, a patient was so convinced that his stepfather was a robot that he decapitated him in order to look for the batteries inside the head (Blount, 1986). To show their commitment, deluded patients often behave in ways coherent with their apparent thoughts. It has been observed that "other Capgras patients have lodged complaints with the police for the abduction of their spouse" (Bayne and Pacherie, 2004, p.6), therefore acting in a way in which there is no doubt that the utterances are meant.

Argument 4) is that delusional patients are usually aware that their statements sound implausible to their audience, and therefore, the patient has not lost the grip of the meaning of what he is trying to say.

In summary, Bayne and Pacherie's claims about meaning do not necessarily attack the root of the problem. As we have seen, their defence of meaning in delusions comes mainly from the fact that patients act and assert as if they mean, literally, what they say. However, these specific types of arguments do not necessarily undermine Campbell's claims. Given that the Davidsonian principle of rationality guides Campbell's assertions, he maintains that the relation between rationality and meaning should be preserved as much as possible in the interpretation. Bayne and Pacherie respect the relation between meaning and rationality too; they disagree only with Campbell's assertion that rationality should be checked in canonical ways. This can be seen from their claims, which take the form of "people with delusions act and speak as if they meant what they say". This form of charity is characteristic of a Davidsonian type of rationality. Their main objection is that meaning should be "bracketed" in order to make sense.

What is the fundamental disagreement between Bayne and Pacherie, and Campbell, about meaning? Campbell argues that Capgras delusion might be thought to involve the content "that woman is not the remembered woman", but, he continues, one cannot ascribe such a specific meaning because the subject does not do the testing that normal people do when in that state. On the other hand, Bayne and Pacherie differ. They think that, in a number of cases, the Capgras patient is sufficiently like a person who holds such a belief and acts accordingly. Therefore, their disagreement is about how much actual patients are like normal people.

What can be concluded in this case from both accounts? One possibility is that the empirical data seems to permit both interpretations. Campbell's account implies that

we do not know what the content of the delusion is. Bayne and Pacherie's account comes with the implication that they downplay what is odd about a person holding such a belief. Nevertheless, both accounts of why the patient does not engage in canonical checking seem narrow.

d) Objections to Campbell's positive account of delusions as framework propositions

i. Bayne and Pacherie's objections against Campbell's positive account

Bayne and Pacherie have several objections against Campbell's positive model of delusions. Campbell argues that the correct aetiology of delusions is a top-down account. Bayne and Pacherie find this surprising, because Campbell claims that the content of the experience does not seem to justify the belief, but in the rationalistic view the subject has no reasons at all (not even poor reasons) to maintain her delusional belief. In Campbell's account, the strange thought seems to be caused directly by organic malfunction. Here Bayne and Pacherie make the following useful claim to argue against that:

> Note that Campbell need not – and probably should not – be read as thinking of the relationship between the brain damage and the belief as causal. The particular way in which a rationalist conceives of this relation will depend on their metaphysics of mental states; one could think of this relationship in terms of supervenience rather than causation. (Bayne and Pacherie, 2004, p.11)

Bayne and Pacherie observe that, in Campbell's view, the justification of beliefs is null compared to the empiricist account: "One might think that acquiring and maintaining a belief without any reasons is more irrational than acquiring and maintaining it for poor reasons" (Bayne and Pacherie, 2004, p.8). They also argue that a rationalistic account does not explain the specificity of monothematic delusions, neither does it explain away why is there some sort of damage to the autonomic system, as one finds in the Capgras and Cotard delusions.

In his model of delusions as framework propositions, Campbell ascribes to them a privileged epistemic status, since one of their characteristics is that they are not subject to empirical scrutiny (Campbell, 2001, p. 96). He also affirms that the semantic meaning of the words used in a (deluded) framework proposition is different from the one an ordinary person would assign to it. As an objection to the former epistemic affirmation, Bayne and Pacherie give as a counterexample cases where the patients have the insight that their delusional idea might be "unbelievable, absurd, an indication that you had gone mad" (Young, 1998, p.37). If patients have this kind of understanding, this means they are questioning the belief, and so, it is not epistemically imperturbable as Campbell affirms. Another counterexample comes from the fact that delusional beliefs in monothematic delusions are highly "encapsulated", compartmentalized, whilst framework propositions are designed to fulfil another role:

> The ability of a delusional belief to function as a framework proposition is dependent, one might think, on it being integrated into the patient's practical and theoretical behaviour. (Bayne and Pacherie, 2004, p.9)

Therefore, as beliefs which are not integrated in the net of the patient's belief, these cannot work as framework propositions, according to Bayne and Pacherie. Campbell also defends a semantic view, in which the meaning of the words of the patient are different (as in a paradigm change). Bayne and Pacherie question why Campbell thinks the latter; it is not the case that the patient *literally* does not know the meaning of his utterances because the deluded patient knows what the term "wife" means. It seems that Campbell thinks that the patient disregards meaning only in the context of his delusional utterance. However, Bayne and Pacherie argue that the latter would contradict Campbell himself, because it would go against his holistic version of framework beliefs. Also, there does not seem to be a better interpretation of what the patient means, other than that he means literally what he is saying:

> As far as we can see, there is no other translation of what the Capgras patient says according to which everything falls into place. Describing the Capgras patient's belief as a framework proposition fails to advance our understanding of what he believes or why he believes it. (Bayne and Pacherie, 2004, p.10).

ii. Thornton's objections to Campbell's view of delusions as framework propositions

In his (2008) paper, Tim Thornton critiques Campbell's proposal that delusions have the same role as framework propositions do. In order to do so, Thornton first observes that Campbell's account seems promising at first glance.

> Firstly, framework propositions provide the context for the practice of giving and asking for reasons which corresponds to making claims of knowledge or doubt. They themselves comprise a background of certainties that are held immune from doubt. They are not themselves objects for possible knowledge but instead make knowledge claims possible. By construing delusions as framework propositions one can begin to explain their incorrigibility and the role they might play in structuring other claims. (Thornton, 2008, p.164)

About Campbell's claim that the true form of the Capgras delusion is "that *perceived* person is not my wife", Thornton observes that framework propositions can be of everyday matters and do not need to be necessarily about perceptions (or sense data), "thus the specificity of delusions is no objection to them playing just this role" (Thornton, 2008, p.164).

Campbell seems to suggest that, once we assume that the patient's utterance is a framework proposition, we can predict that it will be held as a certainty and will not be empirically tested; however, further predictions of the patient will not be reliable because this proposition is a deviant one (and so, it is not shared by other members of the community). Thornton argues that, despite the promising panorama that Campbell offers, seeing delusions as framework propositions does not help us to gain understanding of them. He offers three inter-related reasons for this:

1. Framework propositions are identified from 'within' a worldview.
2. There is a close relation between the epistemic role of framework propositions emphasised by Campbell and Eilan (2000) and a sense-giving role.
3. There are no 'external' criteria for framework propositions.

In the following section I follow Thornton in his explanation of these objections.

1) *Framework propositions are identified from 'within' a worldview.*

Thornton begins by noting that framework propositions are not a natural, homogeneous, unified collection because they are of very different kinds, such as personal or interpersonal; complex or basic; general or particular. This means there is not a unique way to identify them. Also, one has to take into account that they are fundamental, but not foundational, because they are discovered after, not before, the problems that they serve to justify. Therefore, Thornton says "They have instead to be identified piecemeal from "within" a mature worldview" (Thornton, 2008, p.167). Wittgenstein writes about the nature of framework propositions in the following way:

> I do not explicitly learn the propositions that stand fast for me. I can discover them subsequently like the axis around which a body rotates. This axis is not fixed in the sense that anything holds it fast, but the movement around it determines its immobility. (Wittgenstein, 1969, §152)

So, how can one identify framework propositions? According to Thornton, there are two categories: first person and third person framework propositions. In the first person case propositions are identified through stating what propositions need to be justified and which ones are to be held without question. The latter ones will be the framework propositions. However, Thornton observes that this requires a *presupposed* understanding: an empathy about what it is being said (where the word empathy here is taken in Jaspers' sense). In the other case, the third person case, propositions are identified via showing them, that is, they are shown in the linguistic practice of other people: what they say, and what they do as well. However, it is unclear how Thornton would maintain the latter. By looking into a person's linguistic practice one cannot say what his mental state is, and even less, what counts as knowledge or as a certainty. It could be that the person is not aware of his environment, even if the empirical fact seems obvious. The point is that there is no epistemic guarantee to affirm what Thornton says.

But again, even if the latter were the case, this implies that there must be a shared understanding from both parts in order to identify the framework proposition as such. Finally, Thornton summarizes, in order to take something as a framework proposition,

one must first be able to understand it and to recognize its relations to other things that are held as certain.

2) *There is a close relation between the epistemic role of framework propositions emphasised by Campbell and Eilan (2000) and a sense-giving role.*

Thornton considers that framework propositions constitute "the limits of sense" (Thornton, 2008, p.168). Framework propositions describe the techniques of empirical knowledge: they are pre-epistemic, as McGinn describes:

> Thus, my certainty regarding, say, the judgement 'This is a hand' is to be seen as a pre-epistemic attitude that is in part constitutive of my practical ability to speak the language. The judgement that this is a hand is not a piece of knowledge – a true, justified belief, based on evidence – but an authoritative expression of my established mastery of English. (McGinn, 1989, 144)

However, one important issue to bear in mind is that, unlike grammatical rules, the utterance that describes a framework proposition can also describe an empirical claim, given other circumstances[20]:

> But wouldn't one have to say then, that there is no sharp boundary between propositions of logic and empirical propositions? The lack of sharpness is that of the boundary between rule and empirical proposition. (Wittgenstein 1969 §314–9)

Thornton argues that, even if this were case, if we want to see delusions as framework propositions, the delusions have to be, precisely, acting as framework propositions and not just as mere empirical claims. In any case, Campbell is proposing a deviant, or abnormal, framework proposition.

[20] Circumstances in which a framework proposition can change its role and become an empirical sentence might occur. Sentences such as "this is a hand" might point to a fact in the world (as in an autopsy room), rather than asserting that one has the certainty of the fact that this is a hand when pointing to my own hand. But then they stop being framework propositions in such cases.

There might be two ways to cash out the idea of a deviant framework proposition within Campbell's account. However, according to Thornton, neither really holds. The first way to do this would be to propose that there can be a different worldview, with a deviant set of background certainties and knowledge claims, which would be different to our own and that, in addition, cannot be translated. However, the latter would require to contest or refute the arguments of Davidson and McDowell, that is, arguments against the idea that a conceptual scheme divides up a neutral world (Davidson, 1984, pp.183–98; McDowell, 1994).

The other possibility of having a deviant framework proposition consists in postulating a *local* breakdown of the shared background of certainties in some framework propositions. It would be this breakdown that would explain the strangeness of the delusion, if this were the case. However, the important objection would be that this proposition becomes nonsensical:

> But it also suggests that the abnormal propositions both have a kind of sense, because they are framework propositions, but of a nonsensical kind because they cannot be accommodated within a coherent worldview. (Thornton, 2008, p.171)

The latter would make them unsuitable as a framework proposition. However, Thornton says, one could argue that a sense could be given to them. This is similar, Thornton says, to Wittgenstein's discussion about the sense of grammar rules:

> When a sentence is called senseless, it is not as it were its sense that is senseless. But a combination of words is being excluded from the language, withdrawn from circulation. (Wittgenstein 1953 §500)

> When I say that the orders "Bring me sugar" and "Bring me milk" make sense, but not the combination "Milk me sugar," that does not mean that the utterance of this combination of words has no effect. And if its effect is that the other person stares at me and gapes, I don't on that account call it the order to stare and gape, even if that was precisely the effect that I wanted to produce. (Wittgenstein 1953 §498)

Thornton explains: the combination of words “Milk me sugar” has no meaning, but it could be given one in a number of ways. One way could be to change the meanings of the individual words so that the phrase actually “means” bring me sugar. Another way to give it some sort of meaning could be to give to the entire phrase an arbitrary meaning independently of its structure. Thornton says that, whatever the case, this would not capture the strangeness of delusions. In order to do so, words would necessarily have to maintain their standard meaning, their structure of grammar would have to be normal, but what the patient says would have to have no meaning for us. With this constraint, locally abnormal grammar “does not amount to a thinkable thought” (Thornton, 2008, p.171) and this undermines the idea of a deviant framework proposition in this way.

3) *There are no ‘external’ criteria for framework propositions.*

There is also the possibility for Campbell to take framework propositions merely as some beliefs that are held in place whilst other thoughts “turned about” them. This would open the possibility of a description of (delusional) thought independent of its meaning or content. Framework propositions are characterized by not being propositions that one can actually claim: rather, these propositions are expressed by our actions.

> Why do I not satisfy myself that I have two feet when I want to get up from a chair? There is no why. I simply don’t. This is how I act. (Wittgenstein, 1969, §148)

This means that, given the fact that framework propositions have this sort of pre-epistemic and non-psychological characteristics, one cannot describe them in mere syntactic terms and therefore they cannot shed light on delusions as Campbell suggests, nor in any other way. An external account of framework propositions would have to describe a structure of thoughts without identifying the thoughts via their contents. So it would have to claim that something had the form of a framework proposition because of the structure of its inferential role without identifying its content. That faces two objections. 1) It presupposes the possibility of the reduction of meaning to isomorphic structural entities in the head which may be impossible. Crucially, it is not a position a

Davidsonian can adopt and hence Campbell cannot. 2) Some Wittgensteinian framework propositions are more like behavioural dispositions rather than thoughts. I will explain the latter in a different way. Campbell's proposal is to take framework propositions as an expression void of content (because, according to him, their meaning is different and inaccessible to us as interpreters) in which they will still have the role of influencing other beliefs, in a way such framework beliefs will not be questioned. Thornton's objection is that, because framework propositions cannot be expressed, just shown in our everyday actions, they are not epistemic because they cannot say anything about our knowledge and they are not psychological because they say nothing of our mental states either, and therefore, they cannot act even act as "central" beliefs without content.

iii. Thornton's non-positive account of delusion

Many of the accounts of delusions are designed to explain the fact that delusions are strange, but partially understandable; that is, they try to "solve simultaneously for understanding and utter strangeness" (Eilan, 2000). However, Thornton is sceptical that this aim can be achieved, especially through Campbell's account. He thinks both aspects cannot be put together in a satisfactory way: "Interpreting or understanding, but still finding utter strangeness are incompatible goals" (Thornton, 2008, p.173).

To Thornton, a breakdown in the structure of framework propositions invalidates the rational intelligibility of the whole. Therefore, the fact that something makes sense in the first place, implies that one can have (or actually, has) a set of non-deviant framework propositions.

At the end of the paper, Thornton concludes that framework propositions do not make delusions understandable; they can only be *used to understand the fact that one cannot understand them*: "If so then framework propositions are better deployed not as an aid to understanding the idea of delusions but rather as part of an explanation of why they are not understandable" (Thornton, 2008, p.173). In this sense, Thornton's account of delusions is not positive, that is, it doesn't say what delusions are ontologically. It only suggests how the very idea of having abnormal framework propositions does not make sense to us, and therefore, it does not make sense at all:

> So, for the proposal to work – for it to enable us to have some understanding of delusions as a whole by thinking of them as abnormal framework propositions – we need to be able to understand the idea of a genuinely different framework proposition. It is not clear that we can. (Thornton, 2008, p.170)

Moreover, because having framework propositions is also a precondition for something to make sense, Thornton falls into a more pessimistic Wittgensteinian position that it is not even thinkable that a thinker does not have framework propositions. Let us note that this does not imply that delusional people are not thinkers, it just implies that deviant framework propositions do not make sense to any logical thinker.

However, Thornton does imply that there might be a middle-ground position between his non-positive view and Campbell's suggestion of framework propositions. Stanghellini (2004) suggests that, although we cannot understand the idea of a deviant framework proposition, one can propose that a pre-condition of having knowledge claims is to have a definite background as a certainty.

> Without a fully functioning Background, the distinctions between imaginary and real (mind and world), self and non-self, and cause and coincidence fail to be adequately drawn within experience. Also, grasping the meaning of others' behaviours and a weakening of contextual influences are entailed by an erosion of Background sensibilities. (Stanghellini, 2009, p.311)

Thornton then allows such background to have a certain failure: "And thus we can see that such a worldview is vulnerable to failures in that background" (Thornton, 2008, p.174). But this latter idea surely contradicts Thornton's own account, because he proposed earlier on that the precondition for having a framework belief is that it makes sense, and therefore, failures are not actually allowed. Therefore, Thornton has no right to propose the idea of there being a middle ground at all, specifically in this case, Stanghellini's account. Given his own framework Thornton can provide no insight into the nature of delusions this way.

As a conclusion to this section, one might think that Campbell's account is unsatisfactory due to the fact that its own assumptions of rationality and meaning render his positive

view problematic. Campbell's paper has a negative and positive aspect. In the negative phase, he draws attention to the problem of interpreting utterances which express Capgras if one holds to a Davidsonian link between meaning and rationality. In the positive phase, he suggests that rationality can be preserved by thinking of delusions as deviant framework propositions. But because they are, precisely, deviant or non-shared it turns out that we cannot, after all, provide a rationality-preserving interpretation. So we cannot say what the deviant framework proposition is. And that puts under strain the very idea of a deviant framework proposition for anyone already committed to the Davidsonian link between meaning and rationality. However, the question of rationality and meaning is still a valid one, and therefore, it is worth seeing it from a different point of view. To do the latter, in the next section I present an account of delusions that deviates from Campbell's point of view: Sass' account of delusions as a form of solipsism. However, both authors have striking similarities, which I suggest in the general conclusions of this chapter.

IV. Sass' view of delusions as solipsism

a) Sass' positive account

In this section I present an analysis of one of the more influential accounts of delusion that takes interpreting the meaning of delusions seriously. Part of the attraction of this account is the unique way in which analytical philosophy and phenomenology come together in a non-aetiological explanation. In *Paradoxes of Delusion*, Louis Sass (1994) has a very clear agenda: The use of phenomenology and some of Wittgenstein's philosophical concepts to shed light on the meaning of schizophrenic delusions. His approach has some similarities to Campbell's account. To start with, both use Wittgensteinian concepts to construct their positive philosophy. Both are rationalists, as opposed to empiricists, and attempt, further, to argue that the delusional person is, in principle, broadly rational. Both are adherents to the constitutive principle of rationality. I will argue, further, that although Sass, like Campbell, takes the difficulty in understanding the meanings of delusions to be significant, he is no more successful in offering an understandable analysis of them.

Sass states his aim clearly at the start of the book, that in contrast to Jaspers, there is a way out of Jaspers' pessimism about delusions:

> In this book I attempt to do what, according to Jaspers, cannot be done: to comprehend both empathically and conceptually some of the most bizarre and mysterious symptoms of schizophrenia. (Sass, 1994, p.6)

Sass uses the case of a patient, called Schreber, to justify his claims. This patient is famous for having written a very comprehensive and intuitive journal about his own delusional experiences, "*Memoirs of my Nervous Illness",* published first in (1903).
Sass describes Schreber's delusions stressing that they are not due to poor reality testing, nor a failure in logic. Sass shows this by providing some examples in Schreber's writings:

> When the rays approach, my breast gives the impression of a pretty well-developed female bosom; this phenomenon can be seen by anybody who wants to observe me... A brief glance however would not suffice, with his own eyes the observer would have to go to the trouble of spending 10 or 15 minutes near me. In that way anybody would notice the periodic swelling and diminution of my bosom. Naturally hairs remain under my arms and on my chest; these are by the way sparse in my case; my nipples also remain small as in the male sex. Notwithstanding, I venture to assert flatly that anybody who sees me standing in front of a mirror with the upper part of my body naked would get the undoubted impression of a female trunk-especially when the illusion is strengthened by some feminine adornments. (Schreber, 1903, p. 207)

According to Sass, Schreber's own report only states that his body *gives the impression* of having changed and that its feminity is only an illusion. Schreber's delusions are not asserted literally. Complementing the latter, Sass observes that Schreber is not prone to act on his delusions: "[M]y so-called delusions are concerned solely with God and the beyond; they can therefore never in any way influence my behavior in any worldly matter" (Schreber, 1903, p.301-302). This defies the analysis that Schreber's clinician

made of him, where he states that Schreber believed firmly in his delusions and had real hallucinations.

Schreber also had delusions in which he took external world objects to be unreal. In his journal he describes how he is calmly sitting in a park, and then, out of nowhere, wasps appear before his eyes *ex nihilo*. The interesting part, Sass states, is how Schreber starts considering that objects actually exist only when one is looking at them (Cfr. Sass, 1994a, p.81).

Such ways of doubting reality, without asserting that the external world has changed except for oneself, Sass says are strikingly similar to the philosophical doctrine of solipsism, defined as:

> [T]he philosophical doctrine that claims that the whole of reality, including the external world and other persons, is but a representation appearing to a single, individual self, namely, the self of the philosopher who holds the doctrine. (Sass, 1994a, p.81)

Sass recognizes that using the term “solipsism” to describe the patient’s attitude could be misleading because patients are not specifically aware of being engaged in that philosophy. Therefore, he decides to call it “quasi-solipsism” (Sass, 1994, p.39). I will discuss this terminology later.

Having said that, Sass adheres specifically to Wittgenstein’s notion of solipsism even though Wittgenstein was not concerned with proving the truth or falsity of this doctrine; he was trying to dissolve such metaphysical questions. One method of achieving this is to consider what the experiential counterpart of solipsism would be. Wittgenstein said that solipsism can only be achieved through inactivity:

> To get clear about philosophical problems, it is useful to become conscious of the apparently unimportant details of the particular situation in which we are inclined to make a certain metaphysical assertion. Thus we may be tempted to say "Only this is really seen" when we stare at unchanging surroundings, whereas we may not at all be tempted to say this when we look about us while walking. (Wittgenstein, 1958, p. 66)

And through an abnormal intensity of looking at things:

1. The phenomenon of staring is closely bound up with the whole puzzle of solipsism. (Wittgenstein, 1968, p. 309)

2. Ask yourself: what does the word "feeling," or still better "experience," make you concentrate on? What is it like to concentrate on experience? If I try to do this I, e.g., open my eyes particularly wide and stare. (Wittgenstein, 1968, p. 315)

Both are consistent, according to Sass, with the delusions mentioned above. More explicitly, Schreber, in his journal, describes how he remained motionless for hours and in hyper-concentration.

Sass then proceeds with his justification of why solipsism matters in his account: “A number of the characteristics of the schizophrenic delusional world that Jaspers considers inexplicable can, in fact, be comprehended as aspects of quasi-solipsism.” (Sass, 1994 a, p.86). In the first place, solipsism can account for the supposed “incorrigibility” that Sass says is present in Jasperian accounts. To support this, Sass observes that “solipsistic claims cannot be adjudicated in any public domain” (Sass, 1994 a, p.86). Claims, such as one is “represented as” a female, cannot be corroborated by any confirmatory or disconfirmatory evidence. This is an example of what Wittgenstein considers as “seeing as”. In the *Philosophical Investigations* (1953) he states that “seeing as” is a hybrid phenomenon that can be described as a conceptual or a perceptual process (Cfr. Wittgenstein, 1953, &197 &203).

Wittgenstein also remarks that seeing can be affected (for example, in the famous image of the duck-rabbit). But Sass argues that this is not because of a mere act of imagination: “Yet it would be equally mistaken to think of an instance of seeing-as as being merely interpretative or imaginary, thereby assimilating it to a conceptual process” (Sass, 1994, p.31). Seeing-as has some perceptual qualities too, the relevance to delusion is in the interpretation of those perceptual qualities descriptions, as in Schreber’s.

Sass suggests that the phenomenon of double bookkeeping, that is, having two contradictory sets of ideas and endorsing them, can also be understood by analysing Schreber’s experience. He states that there are two “realms of experience” that differ greatly in their ontological status. The first realm is the objective, normal world whilst

the other one can only be seen “in the mind’s eye” according to several of Schreber’s annotations. In his deviant world the agreement of other people is unexpected and irrelevant. His certainty, as is the case with many cases of delusion, is based on his own (first person) impressions, which are indubitable. Schreber states in his journal that, although others imply that his delusions are only due to a vivid, pathological imagination, the others cannot have access to his own experience: “What can be more definite for a human being than what he has lived through and felt on his own body?” (Schreber, 1903, p.132). However, Sass observes that in a solipsistic experience it is not necessary to act because all acts seem unreal and inconsequential.

To a person who believes in a shared, certain, intelligible world, schizophrenic phenomena must seem unreal, Sass concludes, unless he is in a deep solipsistic position, where everything seems to depend on his consciousness; in this case any phenomenon is as real as it can be:

> The distinctive feature of Schreber's lived-world, then, can be described as his failure to accept this premise of a shared, or organizing horizon; and the mark of what we might call his autism is the fact that (like a solipsistic philosopher) he puts his ultimate faith in his own immediate experience, particularly in the experiences he has while in a passive, detached, and isolated position. (Sass, 1994a, p. 90)

Therefore, for Sass, solipsism is a way of explaining the strangeness of delusion comparing it to a not so puzzling philosophical doctrine in which the experience that matters is the one “felt” by the person and where no external agreement is needed. Here, Sass is trying to balance the fact that solipsism is an idea which we can empathically and rationally understand with the other fact that it is also at the same time sufficiently strange and nonsensical that it still is able to capture the nature of delusion. In this sense, Sass is trying to find the delusional sense as a failure *within* rationality, not outside rationality. What Sass is ultimately doing is some sort of *interpretation* of the phenomenon, that is, to consider delusions as making sense within our rational, external point of view.

Finally there are several things one can examine from Sass’ account that I have described in this section. The first one concerns his own labelling of the patient’s account as “quasi-

solipsism" rather than solipsism. It might seem that to make the term softer is a good move because it can be postulated that "true" solipsism requires awareness of actively engaging in this philosophical mood. However, the question of what the term "quasi-solipsism" means arises in its place. Sass coined the term to describe Schreber's and other analogous delusional attitudes. But one has to observe that it only has the independent meaning of "being similar" to solipsism in the way that delusions are "similar" to it. But then its definition has become circular: Delusions are quasi solipsistic because quasi solipsism is like delusions. Therefore Sass fails to give an independent definition of his main idea.

Another question is whether Sass' account is coherent with itself. This question might take a slightly different form: Is quasi-solipsism compatible with delusions? First I will analyse it as Sass would. If we think that solipsism has similarities to delusions, that point is similar to Reimer's doxastic position in which delusions are compared to nihilistic philosophies. It seems that thinking that delusions are akin to philosophy in its different forms is a very tempting idea. It is not completely surprising that Sass' account has something in common with a doxastic account, given the fact that Sass' adherence to the constitutive principle of rationality agrees with some doxastic accounts. However, given the fact that the definition of quasi-solipsism is not independently defined, one must conclude that Sass' account is coherent only because he has defined an ad hoc term to describe the phenomenon: that is, delusions are a sui generis phenomenon with some structural characteristics similar to solipsism. No doubt, Sass would not agree with this definition of his account, and could argue that what he means by solipsism is what Wittgenstein said solipsism was. This, however, takes us to the analysis of the following objections to Sass.

b) Objections to Sass

Rupert Read, in his (2003) response to Sass, agrees with Sass in a common assumption: that any scientific or quasi-scientific explanation of delusions does not provide a better *understanding* of schizophrenic delusions.

> Understanding some of the causes behind schizophrenia, as probably we increasingly do, is not then, in my sense, understanding schizophrenia,

> understanding the people who suffer from it, or understanding the form and flow of their thoughts and life. (Read, 2003, p.116)

However, contrary to Sass' aim, Read maintains a pessimistic position where there cannot be any sort of understanding of people with severe delusions through a Wittgensteinian framework:

> [A] properly Wittgensteinian approach would, I suspect, show that, except in some very remote and vague sense of understanding, there probably cannot be any such thing as understanding the words, actions, and experiences of the very severely mentally ill, those who might perhaps truly be worth calling deeply different from ourselves. (Read, 2003, p.116)

Read considers the case of meaning in dreams. To Wittgenstein, veridical accounts of dreams can only be given outside of the dream context. Assertions such as "I am dreaming" cannot be evaluated as truth or otherwise: one is not entitled to make this claim. As Read remarks, an account of a dream can only be given outside of the dream (therefore inside the everyday world). However, for the schizophrenic's case the latter cannot apply because there is no outside of the psychosis. It is worth mentioning that, unlike other disorders, the current consensus is that schizophrenia is treatable but not curable "So far as we know, although treatable, the illness [schizophrenia] is not curable; everyone who has a schizophrenic breakdown remains vulnerable to further attacks (…)" (Howe, 1995, p.6). In other psychotic disorders with delusions it is relatively common that patients might achieve a complete recovery (for example, in drug-induced psychosis). In this particular debate, Read takes schizophrenia to be a counterexample to Sass' views.Therefore, for Read, an internal account of schizophrenia would not count as testimony, as is the case in dreams. There cannot be any authoritative first-person account, so, such candidate accounts are non-sense.

One can see that the patient's assertions are not commensurable with ours, Read says. In so far as one translates the schizophrenic utterances into our own language one takes away their literalness in order to give them some approximate meaning. Read suggests that it is not the case, as Sass supposes, that one can understand and translate what is ontologically different:

> [W]e do not get any genuine understanding of something ontological (…) through treating it as translatable, as ontic, as optimistic psychologists and psychiatrists tend to. The interpreting into terms that we understand (or translating) of ontologically different language, language that is other than sensical communicative discourse, is just not a good idea. (Read, 2003, p.121)

Therefore, Read says, it follows from Wittgenstein and Sass' arguments that the patients are not engaged any more in the 'game' of reality-testing. And this negative claim is the closest one can get to characterize the game the patients are playing. Therefore, Read's suggestion is that Sass is fundamentally wrong from the very aim of his account because there cannot be any understanding of schizophrenic cases. Also, there cannot be misunderstanding of delusions because they are not candidates for understanding at all:

> I claim that the kinds of resources we humans have for understanding one another—for understanding one another's actions and being, resources drawn upon in literature and elaborated and stylized in the human/social sciences—are largely not present in hard cases of schizophrenia, including in those who (like Schreber) seem shot through with thinking and introspection. (Read, 2003, p.119)

Read's arguments against Sass are similar to Thornton's arguments against Campbell: both think that delusions cannot be compared to something akin to our everyday meaning. To Read and Thornton the deluded utterances are much more similar to a private language that has merely the illusion of sense, but such a comparison leads to no understanding at all. Read makes this position very clear in his (2003) conclusion:

> But most serious schizophrenia does not fall under that heading. It is better seen as the persistent semblance of another language—very much like the semblance of another language that we find in Wittgenstein's private linguist, a philosopher subject to an illusion of sense, an illusion that his words, in the way he finds himself wishing to employ them, mean anything at all. (Read, 2003, p.123)

Sass' general objection to Read's criticism consists in observing that what can be understood, such as solipsism in schizophrenia, are not matters that can be purely seen as black or white. Read's position, according to Sass, is too extreme in the following sense: He follows the metaphilosophy of the new Wittgensteinians that there are no ineffable truths and what does not make clear sense cannot have any sort of sense at all. Sass summarizes Read's position:

> Patients like Schreber and Renée cannot really be understood, just as we cannot really understand logically alien thought— which, in fact, is not really thought at all; in both cases, there is only the illusion of meaning, only the mirage of an alternative mode of thinking or of living. This, at least, is Read's new-Wittgensteinian claim. (Sass, 2003, p.126)

The principle behind Read's interpretation, Sass says, is Davidson's principle of charity, which has always been problematic when it comes to cases that are not wholly rational or logical at first sight:

> Read's application of this new-Wittgensteinian view does highlight a fascinating interpretative dilemma that has special significance for psychopathology. A key assumption of hermeneutical interpretation is the so-called principle of charity— the assumption that (other things being equal) one should consider the perspective one is attempting to understand to be as rational and coherent as possible, and, indeed, that the finding of coherence actually functions as one criterion (not an absolute criterion, of course) that argues for the likely validity of one's interpretation. (Sass, 2003, p. 126-127)

That seems like a strange criticism coming from Sass given that Sass himself adheres to rationality principles: Sass thinks that delusions are situated on the edge of rationality. However, Sass thinks that it has its limits:

> Elsewhere I have argued that if the principle of charity is overemphasized or applied too rigidly, it can lead to oversimplified interpretations of a philosopher's viewpoint, interpretations that fail to capture ways in which that philosopher

> may be responding to insurmountable tensions within human knowledge or existence. Inconsistency, even contradiction, may sometimes be the price of addressing certain kinds of philosophical problems. (Sass, 2003, p. 127)

Therefore, Sass does not agree with Read's application of the principle, although he agrees with a Davidsonian framework of rationality.

Sass concedes that his solipsistic, or quasi-solipsistic interpretation has contradictions as well. However, the contradictions are surmountable; they only reflect the fact that solipsism is also a contradictory doctrine, its sense is not a logical, Davidsonian sense: "Indeed, the advantage of Wittgenstein's analysis of solipsism is that it allows one to capture, in fairly precise terms, the nature of the solipsist's equivocations and contradictory tendencies" (Sass, 2003, p.128).

Read's response to Sass' objections is that philosophical positions that contain "inexorable tensions", such as solipsism, cannot be really understood (Read, 2003a). To Read, trying to represent what is ineffably enigmatic can only result in fatally undermining the position they are trying to clarify. Solipsism cannot be the answer to anything, according to Read, because it is itself a "delusion of sense" according to Wittgenstein:

> Wittgenstein's great achievement, in wonderful therapeutic detail in his later work, was to show that we do not have a clear understanding of this [solipsism]; or rather, to show that there is no *it* here. (Read, 2003a, p.137)

Read also argues that solipsism cannot be understood, only represented or mimicked, and concepts employed by Sass to gain empathic understanding, such as "hyper-rationality" and "hyper-reflexivity" are not comprehensible and hence amount to nothing. Related to this point, Read disagrees with Sass' fundamental point that delusions are to be *interpreted* (under the light of solipsism). To Read, interpretation leads to distortion of the phenomenon. The phenomenon should be *displayed* as it is, with their utterances and behavioural consequences, and judged under this light.

Sass also objects to Read's assertion of Jaspers-like incomprehensibility applied as a generality to all mental illnesses. He argues that some of the milder mental diseases can be perfectly understandable, and he observes that most people will deviate from perfect

logical thinking and that he seems to imply there is a definite line when one is not a thinker any more:

> But if we recognize that the less than "very severely mentally ill" or the insufficiently severe cases are, in significant measure, comprehensible, then it seems odd to talk as if, somehow and somewhere, there is a kind of absolute line beyond which all understanding falters, beyond which all interpretation can be no more than complete and total folly or utter projection. (Sass, 2003, p.130)

However, Read replies that, although he thinks that schizophrenic patients are incomprehensible and their utterances without sense, they still deserve a high degree of compassion.

So, is Sass' account successful within a Wittgensteinian framework? The short answer to this is that it is not successful because, in those terms, his account is nonsensical. Read's comments show why the account does not make sense of delusions. This is because solipsism, itself, is strictly non-sensical. Sass argues (in a personal communication described in Thornton, 2012) that this depends on the interpretation of Wittgenstein that one takes. He maintains that it is only 'New Wittgensteinians' (adherents to the 'resolute' reading of the *Tractatus*) who assert that what is nonsensical simply lacks meaning or sense. The traditional interpretation held that Wittgenstein did think that philosophy could in some way point beyond the limits of sense prompting Frank Ramsey to object 'What we can't say we can't say, and we can't whistle it either' (Ramsay, 1931, p.238). However, this defence suffers two related defects.

First, even if the traditional reading of Wittgenstein is correct, that only explains what Wittgenstein meant to do in the *Tractatus*, to set out somehow, without saying it, how language connects to the world. The interpretation alone does not explain how he could have been successful in this. If, as Thornton's (2012) suggests, Sass' project consists in using a special kind of nonsense to shed light on delusional thoughts, then Sass owes an account of what light is shed. The traditional view looks most plausible in the connection between language, thought, and world through the picture theory of meaning. Sass, however, focuses on solipsism.

Second, the question then is: So, what is the truth in solipsism if solipsism is strictly nonsense? The traditional reading of Wittgenstein accepts that there is a break between

his early and later philosophy, because the New Wittgensteinians argue that they express the same broad views. In his later philosophy, Wittgenstein does connect nonsense with having no sense. So, Sass seems to be committed to the project of defending the earlier Wittgenstein against the later Wittgenstein. However, he has not undertaken this work.

Therefore, in the end, Sass' account is also unable to give a coherent picture of what the meaning in delusions is through his approach of the latter as expressions of solipsism.

V. Conclusion

Berrios, Sass and Campbell address an important, but rather neglected, issue in the philosophy of delusions: they ask whether delusions have content by analysing their meaning. However, their search for meaning comes to an unexpected halt due to their own assumptions about rationality. When one asks a delusional person if they mean what they say they invariably answer that they do. However, for these three authors, rationality constraints dictate that it is not clear what the patient actually means. Even with the hardest of intellectual efforts it is hard to see how it could be otherwise. If the patient were coherent and rational in every aspect, then one would not call him delusional. If the patient had permission to "suspend" rationality and the meaning was obscure only in a single sentence, as Campbell suggests, our normal, shared meaning would not work, as Thornton observes. If the patient were on the border of rationality, as Sass maintains, the border would have to make sense, but as Read highlights, that cannot be the case. However, if one requires too much from the notion of rationality, as Berrios does, one ends up with a notion that cannot encompass everyday beliefs, as Bortolotti observes.

Berrios' account contrasts with Sass' and Campbell's in many aspects. Berrios endorses a bottom-up account, and maintains that patients cannot be assumed to be rational. He specifically denies that delusions have any useful content. Moreover, delusions are not beliefs, not even "wrong beliefs". What the patient says is just a fragment of random crystallized information. On the other hand, Sass and Campbell hold rationalist (as opposed to empiricist) views; they suggest the delusion is a top-down disturbance of the subject's beliefs. Furthermore, Sass and Campbell assume that the patient is broadly

rational, and so, that principles of rationality should prevail in charting the meaning of the delusional state. The latter is shown by Sass' interpretation of Schreber as an agent that is as close as possible to rationality as solipsism would allow, and the fact that Sass thinks delusions are not senseless in a Wittgensteinian way. Campbell's interpretation is close to the constitutive principle of rationality too. But, although Campbell and Sass are very similar in those aspects, Campbell and Berrios have some things in common as well.

Berrios and Campbell strip out the meaning from the delusional utterance. Both of them apply their particular criterion of rationality to delusions and conclude that there is no way of conciliating the utterance (the speech) with any sort of meaning. Campbell and Berrios think that the limits of rationality end at the point where the patient does not pass a specific proof: for Campbell rationality ends when the patient does not engage in canonical testing; for Berrios it ends when his notion of belief is not met.

Sass's account, by contrast, does not agree with Berrios' and Campbell's pessimism about meaning because Sass thinks there can be a way of balancing the fact that delusions apparently do not make sense with the idea that they possess a shared meaning. Sass thinks that interpretation can be achieved by thinking of schizophrenic delusions as a form of solipsism. In his account, solipsism lies on the boundary between two realms: the senseless and the realm of shared meaning. Solipsism lies at the limits of what makes rational sense, although it cannot be taken seriously in practical life. Sass thinks it is the same in the case of delusions.

It is clear that the boundaries where rationality ends are different for the three philosophers, and so their accounts differ in important ways too: to Campbell meaning in delusions is non-commensurable whilst in Sass it is possible to make meaning commensurable by comparing it to solipsism. Berrios' account is more extreme, delusion does not and cannot have any meaning; it is just a random fragment of information because meaning can only be found in the pre-delusional state, before the delusion is patent to others.

Finally, Sass and Campbell are trying to defend the meaning of delusions by proposing they are akin to something familiar, although in both cases, the familiar part is strange: deviant framework propositions and solipsism. They do so in order to preserve the sense in which delusions are bizarre. So, they are trying to address the apparent difficulties in understanding the patients' utterances, however, this leads to the same difficulty. It is

therefore plausible to object that one does not really understand either deviant framework propositions or solipsism in the first place. However, the solution to this problem is not that delusions are empty speech acts either. In this case, the aspects that can be understood of the delusion are lost. The fact that, in the end the three authors in this chapter are unsuccessful in applying rationality principles says something about the phenomenon of delusions: it shows how the contradiction arises, as it does in the rest of the chapters.

CHAPTER VI- AN APPROACH TO DELUSIONS THROUGH THE LIKENESS ARGUMENT

In this chapter I will set out a key reason why there is fundamental and continuing disagreement about the nature of delusion. To do so I will describe a related case in philosophy of mental health in which there is ongoing disagreement: In his book, The metaphor of mental illness (2006), Neil Pickering presented a model which is meant to explain disagreement in the latter debate but which I will use to shed light about delusion. His main claim is that the reason for the disagreement is the form of the arguments offered. I will discuss in this chapter that this holds for the case of delusion as well.

After I present Pickering's analysis of the debate about the status of mental illness, I will defend his account against a number of objections. My aim will be to explain that such criticisms have mainly misunderstood the reach and implications of Pickering's claims.

After that, I will show how accounts of the nature of delusions correspond to a certain form of argument, and that this structure is vulnerable to important objections. I finish with a substantial claim based on the analogy of Pickering's description about the status of the debate: there is a fundamental problem with the philosophy of delusion, and therefore, the debate cannot be settled as it is now, and there is no prospect of it happening anytime soon.

a) Why do we say there is disagreement in the first place?

In this section I will argue that the form of the disagreement in the context of philosophy of delusion is significant, in the sense that it is different from other philosophical disagreements. In principle, it is not entirely surprising that there is some sort of disagreement in any particular philosophical debate. It is not clear, just looking at the fact that there is disagreement that it is, in fact, different from ordinary discrepancy. After all, philosophy works with many different assumptions and part of the richness of the discussion is precisely that there are potentially various points of view to consider.

However, although philosophical debate seems to be characterised by some sort of disagreement, there seems to be something significant about the nature of it regarding delusion. In the end, what differentiates both disagreements relies on the diagnosis I present in the following sections, that indicates that there is a specific problem, both in the debate about the nature of delusion and also in the debate of mental illness as a category. From such diagnosis I conclude that there are good reasons to believe the debate will not be settled soon. In other words, the form of disagreement is significant; this relies on the form of diagnosis I offer in the upcoming sections.

In the previous chapters of this work I highlighted the way that different philosophical accounts of delusion have different theories, which result in different theories of the ontology of delusions. The fact is that there seems to be no agreement among accounts, even if there are many self-coherent and rational ones. For example, the notion that delusions are beliefs generates discussion about what type of anomalous beliefs they are, each with a different characterization. It is noteworthy that in the debate about delusion there is disagreement rather than agreement about the surrounding concepts, for example, the concept of belief. Seen from afar the debate naturally takes us to a difficult position where it is not possible to declare that there is a predominant tendency, nor an account that is "more correct" or "closer to the truth" than others. All the accounts have their weak and strong points, but there is no agreement about which one correctly describes the nature of delusion. In summary, there is not a conclusive analysis of the nature of delusion in the current literature. Following this, it is possible to pose the question 'Why is there no such definitive analysis?'

In the case of delusions the reasons for disagreement seem to be less clear because all the aspects that constitute the concept of delusion are in dispute or have been questioned. Such aspects are, mainly, the concepts of rationality, belief, experience, and meaning. For example, whether or not delusions are doxastic was not questioned until a couple of decades ago because being a belief seemed to be the core of what was intrinsic to being a delusion; however, once it has been brought into question, we do not have a definite answer to the question whether delusions are beliefs or not. A similar case concerns the question of the meaning of particular delusions: whether they mean something (specific, understandable or shareable) is still debated, as Sass (1994) and Campbell (2001) argue. In sum, the aspects of the definition of delusion are in question,

and agreement seems far from reach. So, the question that has been present throughout my work, that is, what is the nature of delusion? has not been answered, and it is an interesting and substantial question why this is the case.

My proposal for explaining the disagreement in the debate about the nature of delusion comes from it sharing a similar form of argument exemplified in a different debate: the debate about the reality of mental illness. In the next section I will explain my diagnosis, which is based on Neil Pickering's analysis of why there is disagreement in the latter debate in order to apply it to shed light on the prospect about there being no agreement in the debate about delusions.

I. Neil Pickering's discussion about the likeness argument

In the previous chapters of this thesis I examined the current debate about the nature of delusion. However, I am going to look at an argument from elsewhere in the philosophy of mental health in order to offer a diagnosis of why there is ongoing and fundamental disagreement about delusions. Many important authors have contributed to the dispute, such as Boorse (1977, 1982), Fulford (1989), Wakefield (1992), Culver and Gert (1982), Brody (1985), Szasz (1960), Flew (1973), Claridge (1992), and Brown (1985), to name some[21]. Even though many decades have passed since this discussion began, there still seems to be no agreement.

What is contested in this debate is whether mental conditions, such as alcoholism or schizophrenia, should be categorized as an illness, moral weakness, or some other concept. This debate consists in stances where mental illness can be accounted for in different ways that do not necessarily imply being an illness as such; for example, some accounts argue for a radically different description, such as being a moral failure.

Neil Pickering (2003) proposes a plausible diagnosis of why this is the case. He argues that agreement in this matter cannot be reached, and this is due to the form of the argument used by the competing philosophers and psychiatrists to justify their positions.

[21] See Pickering (2003) for a brief description of their main claims.

The “likeness argument” is Pickering’s name for a kind of argument used in this debate to try to settle the status of mental illness. According to him, the dispute relies on a specific form of an argument from analogy. Such accounts, Pickering argues, use a specific structure that states that a certain phenomenon should belong to a certain category because it is *like* another phenomenon for which there is no dispute about its categorization.

The likeness argument works as follows in the case of mental illness: certain phenomena are illnesses because they are sufficiently like other conditions that are accepted as an illness. Its main procedure is to highlight the features of the phenomenon in question and compare them to physical illness. Physical illness has historically been considered paradigmatic illness; physical illness is unproblematic as a concept (unlike mental illness), and therefore, is often used as a standard for other more controversial concepts of illness. Therefore, in this debate, if the features are enough like physical illness then the phenomenon is deemed as mental illness. In the case where the proponent wants to show that the phenomenon in question is not an illness, the author will argue that the phenomenon is not an instantiation of a paradigmatic kind of physical illness. The procedure above is summarized as follows:

> So-called mental illnesses will rightly be classified or categorized as illnesses to the extent that these conditions have the features of illnesses; and to the extent that they lack these features, classifying as illnesses will be a mistake. (Pickering, 2003, p.244)

The likeness argument assumes two things:

1) That the features of human conditions such as schizophrenia decide what category, or kind, these conditions are a member of.

2) That, with respect to the presence or absence of such features, a condition such as schizophrenia is describable independently of the category it is assigned to (Pickering, 2003, p.245).

Assumption 1 suggests that there are sets of objective features that determine the categorization of a certain concept. Assumption 2 proposes that it should be possible, in principle, to assert the existence of some features independently of a prior top-down theory. These assumptions are desirable in every account in order to avoid question

begging in the first place and ensuring that the categorization is unbiased. However, Pickering argues that those assertions do not hold true, and therefore, he highlights that the likeness argument fails rationally to persuade and hence there is continued disagreement.

Pickering has two objections to why assumptions 1) and 2) do not hold for the mental illness debate: the weak and the strong objection.

a) The weak objection

The weak objection focuses on the first assumption. It observes that, although mental and physical illnesses have features in common, they also have differences that are being disregarded. Pickering portrays it as follows:

> The weak objection asks why the similarities should be taken to bring these conditions into a single category, when the differences do not have the effect of putting them in different ones (…) The weak objection, then, is that the likeness argument actually relies on a decision to interpret, or see, the likeness argument as having significance. (Pickering, 2003, p.246)

The weak objection thus targets the first assumption. But if that assumption does not hold, then the application of the likeness argument will depend upon interpretations that can rationally be contested by competing parties.

The idea is that values play a role in the motivation for defining a classification but that once the classification is set up, it is the objective properties of things that fix their status. Thus, even if values play a role in marking out the class of illnesses why that class is important to us, that role might be merely in the background. Having agreed what the weak objections says, that is, that in many cases, classification promotes more weight to certain features rather than others, one can argue that this failure is not fatal to agreement or classification itself; one can think that it can be bypassed. This can be the case according to Pickering: if one is able to acknowledge that human decisions play a role in assigning the weight of the features of a certain phenomenon, the weak objection does not have to be undermine the likeness argument.

> This is a significant concession to the likeness argument, and one which I think plausibly strengthens it by showing how to take into account the widely acknowledged role of values in categorizations such as illness without giving up the first assumption. (Pickering, 2003, p.248)

However, the strong objection is unavoidable and sufficiently fatal to any classification system that has the form of the likeness argument, as I will argue in the following section.

b) The strong objection

The strong objection is focused on the second assumption. It argues "that the ascription of features to conditions depends on the overall category (…) into which they are placed" (Thornton, 2007, p.29). Therefore, there is no reason to expect that competing parties will agree on the same description of the features of a condition whose status (as an illness) is in dispute. Hence the likeness argument cannot provide a neutral arbiter.

The central issue about the strong argument is about the role one assigns to the concept of illness. As Pickering shows, there are many philosophical accounts that have incompatible answers to the problem: alcoholism can be seen either as a failure of will or as a need created from brain chemistry even when looking at the same facts. Schizophrenia is seen as a dysfunction of the mind by some, and to others, it is a failure in the normal brain monitoring process. Pickering argues that both views are irreconcilable because it is unclear what sort of function has the task of monitoring thoughts. Therefore, what features are relevant for the description is dependent on the theory one is adhering to in the first place.

It is possible to broaden both the weak and the strong objections in the form of theory ladenness of observation in order to make them more explicit:

-The weak objection*: In order to judge the relevant phenomenon as belonging to a certain category people evaluate the features as having a particular weight according to their own theory.

-The strong objection*: Even the particular framework will influence the features or characteristics assigned to the phenomenon therefore begging the question from the beginning .

I make use of this version of Pickering's objections in several parts of this chapter rather than his original formulation because these claims are a better explanation due to their more general approach.

II. Responses to criticisms of Pickering's arguments against the likeness argument

The purpose of this section is to describe and respond to the main arguments against Pickering's view. Although examining in depth Pickering's arguments may seem to be a digression from the main body of the thesis, I am looking at his analysis of a debate in a different area of philosophy because later in this chapter I argue that the same broad analysis applies in the philosophy of delusions. Therefore, I will begin with some of the more specific criticisms about Pickering's statements and then I will continue with the more general approaches until I address the most significant criticisms. I will argue that, providing one takes note of the context of the likeness arguments that Pickering examines, his diagnosis that an argument of this form cannot rationally resolve the issues remains sound.

a) Pickering's defence against possible objections

In his original paper Pickering anticipates and responds to some of the possible criticisms against his account. He divides these comments into two categories: those that target the weak objection and those against the strong objection.

i. Against the weak objection

Against the weak objection one could advance the following argument: One might think that although there might be some features that are merely accidental, there might be others that are essential in order to belong to a category. Pickering (2003) argues that, although this can be true for some very basic categories, it is not the case for something as complex as illness. If it were the case that there were definite characteristics that

were necessary to belong to a category then there would have to be a one to one relationship between the members of the category and a set of features. However, Pickering observes, that the latter is highly unlikely due to the wide range of things we consider as illnesses. For example, there are cases in which being an illness corresponds with the feeling of being ill, as in pneumonia; but there are cases, like hypomania, that do not correspond to such categorization. Therefore, the one to one relationship does not apply. In the case of mental illnesses, an extra criterion has to be introduced, such as the concept of being a dysfunction, in order to say that both hypomania and pneumonia are illnesses.

This takes us to the next argument: Pickering suggests that, in the case of categorizing illnesses, value judgements necessarily have to take place. To him, categorization is not something that is objectively determined by the phenomenon itself independently from a set of interests or values. In other words, the likenesses or differences cannot determine whether a phenomenon belongs to a category without applying judgement. However he argues that this need not undermine the first assumption:

> Relating all this to the likeness argument, the point is that the role of human decision, interests, and values is taken up in the development of a system of classification of human conditions, including the distinction of conditions into illnesses and non-illnesses. But once that system is developed, the classification takes on a life of its own, reliant on the similarities the system of classification caused to be picked out. (Pickering, 2003, p.248, italics added)

The idea is that values play a role in the motivation for defining a classification, but once the classification is set up, the objective properties of things fix their status. Thus even if values play a role in marking out the class of illnesses, in why that class is important to us, that role might be merely in the background. Having agreed that in many cases classification promotes more weight to certain features rather than others, one can argue that this failure is not fatal to agreement or classification itself; one can think that it can be bypassed with sufficient attention. This can be perhaps the case according to Pickering. However, the strong objection is unavoidable and sufficiently fatal to any classification system that has the form of the likeness argument, as we will see in the following section.

ii. Against the strong objection

Against the strong objection Pickering considers the following arguments: It might seem simply much more natural in the first place to put a certain phenomenon in a category rather than in another. For example, in the case of thought insertion there is, arguably, a natural inclination to categorise it as an illness of some sort, rather than a non-illness. However, Pickering observes that this is only because of our underlying assumptions. If one thinks of a heart as a pump naturally it is because of our way of thinking of the body as a machine in the first place. In the case of thought insertion, Pickering argues that some people might claim that the phenomenon itself would suggest its own classification; to them, thought insertion is a symptom of something clearly going wrong in the patient, and although clever alternatives can be found, this is done with some distinct effort (Cfr. Pickering, 2003, p.252). Pickering argues that, although some phenomena do seem thus and so, this might be for weak motives, for example, historical reasons.

> The interest in the genetic basis of schizophrenia, for example, is historically more or less contemporary with the idea that those mental conditions we might now put down to schizophrenia should be seen as a medical problem. (Pickering, 2003, p.249)
>
> The idea that the heart is a pump seems, likewise, a perfectly natural idea, one that appears irresistibly suggested by what the heart does. But the machine model of the heart, it can be argued, is based on a deeper acceptance that the body as a whole is a machine. Yet this is a view with a historical starting point, which holds sway now only after many years of dispute with alternatives, and which may yet have challenges to meet. (Pickering, 2003, p.252)

A different argument against the strong objection would be accusing Pickering of saying that any description is possible given a particular framework. In other words, someone could argue that given a certain set of theories, all parts of the debate have open-ended resources to create arbitrary accounts. However, Pickering clarifies that this is not a necessary implication. If no correlation is found between some features and the theory, then it does not need to create it. Pickering explains:

> The position is that categorizing alcoholism and schizophrenia as illnesses brings with it a description of these conditions in which the features of illness appear. But this is not the same as saying that any description at all is possible. The observed neurochemical correlation hypothesized in the case of alcoholism is what is described: the description of the correlation may be a causal one, or the correlation may be described as a source of moral temptation. However it is the description of the correlation in causal terms, and not the correlation itself, which is the product of the categorization. (Pickering, 2003, p.252)

Another possible argument against the strong objection is that there is a one to one objective relationship between a set of features and a concept that does not take in account broader theoretical perspectives. Pickering undermines these objections by showing that, in all the attempts to describe mental illness, the relevant features of the category necessarily form part of such description. However, most theories today accept that some sort of judgement is involved in categorization. However, even if that is the case, it does not invalidate Pickering's point that people still use (sometimes consciously) some sort of judgement which ends up being part of the same phenomenon they are trying to categorize. Therefore, the strong argument still maintains its validity.

b) Response to Tyreman

Steve Tyreman (2003) agrees with Pickering regarding the inadequacy of the likeness argument to settle the debate about the nature of mental illness. However, he argues that his account of mental illness as a failure of agency is not affected by Pickering's views because it does not follow the form of the likeness argument. It will be helpful for the later part of this chapter, however, to show that it does and hence falls prey to Pickering's critique.

Tyreman works with Pickering's concept of the likeness argument as if there were two different versions of it, which should not be treated in the same way. The first version is one in which the likeness argument is a way of maintaining that physical illness is paradigmatic and epistemically privileged. The other version of the likeness argument,

according to Tyreman, is a generic account. However, he thinks these views do not fit neatly together:

> It is not clear what Pickering is trying to do here by putting two very different views together; the first, the paradigmatic version, holds that physical illness is the gold standard for determining whether something is or is not an illness. (…) The second variation, the generic, sets out to define illness (…) without basing the definition on any particular form of illness. This definition is then used to establish whether a variety of conditions, including mental illness, can be defined by it. (Tyreman, 2003, p.270)

In the first version, Tyreman argues that there are no principled reasons for the concept of mental illness to be measured against the supposedly paradigmatic concept of physical illness. Tyreman's position becomes clear when he spells out what are the assumptions that he sees in Pickering's paper, that is, that the likeness argument works by comparing in what way the current accounts are similar to the standard concept of physical disease:

> [T]here is an implicit assumption that physical illness is the standard against which all other claims to be illnesses should be tested; that is, that the reality of physical illness is largely unproblematic and capable of being used to test other claims. This latter point is not so much an assumption of the paper per se, but it is an assumption that lies at the heart of the likeness argument. (Tyreman, 2003, p.268)

In his criticisms Tyreman agrees that the likeness arguments fails, but for different reasons than Pickering's. Tyreman's argument against the first, paradigmatic version of the likeness argument is that there is not a valid reason of why physical illness should be privileged in the role of deciding whether mental illness is really an illness. In other words, there is not a valid reason of why physical illness should be the measure of all illness. His argument against the second, generic version is that it is truistic; Tyreman claims that accounts from the generic version, like Donagan's (1978), already presuppose the concept of disease:

> The problem for Pickering's argument is that on the generic account he is left with the truism that illnesses are categorized as illness. That is a perfectly valid

> reason for rejecting the likeness argument or at least the second assumption, but it is not the one that Pickering has given. (Tyreman, 2003, p.27)

It is puzzling how Tyreman then thinks that his own generic account is immune to Pickering's objections to the likeness arguments, even more so given that Tyreman argues that his rivals, who hold that mental illness is indeed an illness, are affected by Pickering's objections. He writes this explicitly towards the end of his paper:

> The point is that the validity of an illness is judged, on the medical model, by the presence or absence of disease and that is where the likeness argument can be shown to fail. (Tyreman, 2003, p.271)

The fact to take in account in order to make sense of the latter is that Tyreman has a very entrenched idea that there are actually two versions of the likeness argument, when this is not actually the case; Pickering designed his account as a criticism to all the versions of the likeness argument that play a role in the debate of mental illness. However, by proposing two versions of the likeness argument, Tyreman thinks that Pickering's objections do not apply to his own generic criterion. Tyreman defends his account (similar to Fulford's) by attacking mostly the paradigmatic account. Therefore, it is clear that Tyreman misinterpreted Pickering's explanation about how the latter's view actually works because he does not see that the objections might apply to him as well. But, as Pickering replies: even if we accept his model of illness as action failure as correct, it still does not settle the issue of whether, for example, alcoholism is an illness because of the strong objection. We beg the question when we describe the alcoholic's weakness of will as an action failure.

> My claim is that it is framing the question about the reality of mental illness in this way that is the problem, as once we ask "is an alcoholic experiencing the failure of agency?" my strong objection to the likeness argument applies. Accounts of alcoholism seem available in which a failure of agency is not present, and describing it in illness terms then seems to presume what it is supposed to prove. (Pickering, 2003, p.273)

In that sense, Tyreman assumed too much (although he acknowledges that such assumption is not explicitly in the paper) and ignored a crucial part of the likeness argument: that it goes both ways in the debate. The paradigmatic version of the

argument is supposed to say that mental illness is illness because it is *like* a paradigmatic, physical, illness. However, what Pickering actually says is *that the form of the argument itself is the problem*. In that sense, Tyreman's account also falls prey to the flaws of the likeness argument.

Summarizing, by assuming that Pickering's criticism has two different versions, Tyreman agrees with Pickering for the wrong reasons; furthermore, Pickering's arguments actually work against Tyreman's account of the reality of mental illness.

c) Response to Radden

Jennifer Radden (2006) wrote a review of Pickering's (2006) book where she doubts the validity of his main argument. Although she praises most of Pickering's efforts, she disagrees with his strong objection. As I have described above, Pickering considers two views for each of the following phenomena: ADHD (Attention Deficit Hyperactivity Disorder), alcoholism and schizophrenia. He argues that two incompatible interpretations can be derived from two different starting assumptions and that such interpretations are determined by the same theory they are arguing for. Radden's criticism is that Pickering's description of mental illness as a non-medical condition is borderline plausible in the cases of ADHD and alcoholism, but impossible to make sense of in the case of schizophrenia. To her, it is only through the language of dysfunction, illness, disorder, etc. that schizophrenia can be described rather than a willing effort from the patient to believe in a distorted reality, as Pickering said. Radden therefore asserts that this lack of a plausible alternative account leads to a way out of the problems of the likeness argument.

In principle it is true that Pickering's example about schizophrenia is not an account that experts or patients would be willing to endorse. But that does not undermine Pickering's case. The alternative, mostly implausible theory about schizophrenia is in the text only to show *possible* interpretations that might appear given the likeness argument. An example Pickering could have used as well is that patients typically deny that there is any underlying condition beneath their particular experiences: they do not think there is something wrong with themselves, they think there is something going on in the world

that affects them (for example, they might think someone is out to get them). With reference to delusions described as morbid (i.e. identified as schizophrenic), Jaspers himself expresses the thought that it is not possible to say, of such experience (especially what he describes as 'metaphysical experiences'), whether it is a pathological experience or a spiritual experience, thus reinforcing the point about the availability of multiple interpretative frameworks, not just as a logical possibility but as a salient one (Cfr. Jaspers, 1913, p.107). Another factor to take in account is that Pickering is implicitly using the notion that science in this field will advance and develop new theories that will have part into the classification of the phenomenon as an illness or not. He is not really arguing that his ad hoc theory of schizophrenia is correct, he is merely exemplifying how a future theory, with different assumptions, can perfectly compete with our current, most accepted theories. As Pickering writes: it might seem natural to see schizophrenia as an illness, and it is also natural to see the heart as a pump, but only under the previous assumption that the body itself as a machine (Cfr. Pickering, 2003, p.252). As I just pointed out, seeing schizophrenia as an illness is contested by Jaspers, therefore, there are other views available which exemplifies that this phenomenon does not have to be considered necessarily as an illness.

There is also Radden's second worry: that Pickering's theory will be obsolete once brain science has determined the aetiology of mental phenomena. In this case I suppose that Pickering's argument would be that even the best theory of how the brain works, cell by cell, would not change at all the debate about unusual mental phenomena as an illness: it would still come down to what the theory predicts in the first place. Because even if brain cells do something very specific, that action would not differentiate an act of will from an act from an illness. An instantiation of this is: if it turns out that the brains of homosexuals were different (and surely they must be in some respect) to heterosexuals, that would not imply that the difference amounted to pathology. So, I don't consider that Radden's claim affects Pickering's discussion about the flaws of the use of the likeness argument.

d) Response to Gipps

Richard Gipps (2003) has some concerns about Pickering's strong objection. Gipps accepts that it might be the case that there is some question begging in the way the debate has been held: "theorists who deploy the likeness argument (…) typically presuppose what both we and they had all along hoped would be demonstrated in their argument" (Gipps, 2003, p.255). However, he raises the following observations about why this is not a problem: a) that there does not have to be an agreement of a certain description of the phenomenon in order to be something to argue about, b) that the validity of the likeness argument should be proved case by case and not as a generality (even within the boundaries of the mental illness debate), c) that some of the features are constitutive of the phenomenon and just *happen* to be in this specific phenomenon.

About a), this is an interpretation that Gipps gives of Pickering's strong objection, which is:

> That, with respect to the presence or absence of [those features of conditions which features decide whether or not the conditions are illnesses], a condition such as schizophrenia [must be] describable independently of the category it is assigned to. (Gipps, 2003, p.255)

Gipps says that, although this seems a reasonable requirement, if this were the case, everyone would have in principle the same premises and everyone would arrive through logical arguments to the same conclusion, therefore, there would not be any interesting philosophical discussion.

The problem with Gipps' concern in this case is that Pickering's strong objection does not imply that everybody should have the same framework or the same premises in order to reach agreement about the debate of mental illness. Pickering's strong objection requires that the premises of the question about the nature of mental illness do not presuppose what the phenomenon is in their description of the phenomenon. In his example about alcoholism, the question whether the latter is a moral failure takes into account only the features that describe alcoholism like a possible moral failure when the fact is that there are other possible descriptions. Rather, Pickering is asking

for objectivity rather than previous agreement. In other words, Gipps is criticising Pickering's claims for something they are not designed to do.

With respect to b), Gipps is trying to undermine the whole of Pickering's objections to the likeness argument by saying that it is too general. Although it is true that Pickering's view can be analysed case by case, he has proved his point with concrete examples that Gipps has not been able to debunk. Whether there might be cases that Pickering has not foreseen, that might be the case, but so far that is not the case.

Regarding c) Gipps argues that some of the phenomena are classified in a certain category due to them being constitutive of our notions of health or illness. In other words, Gipps is suggesting that "[Pickering] remove[s] the one thing we all found too obvious to need saying: someone who is deluded is ill" (Pickering 2003, p.274).

However, Pickering's response is that, even if some ways of thinking about the mind, body, or symptoms are deeply engrained, that does not address the normative point about whether they *should* be; as he points out:

> That something seems natural in this sense need not, however, be taken to be the final word on the matter. The idea that the heart is a pump seems, likewise, a perfectly natural idea, one that appears irresistibly suggested by what the heart does. But the machine model of the heart, it can be argued, is based on a deeper acceptance that the body as a whole is a machine. (Pickering, 2003, p.252)

Pickering argues that it is not "obvious" what is the nature of the phenomenon and what phenomena just "happen" to be part of the description, otherwise there would not be fundamental disagreement in the first place, and even more: this sort of disagreement is precisely what philosophy should investigate; and what is constitutive of a certain notion can change in time (Cfr. Pickering 2003, p.274). Pickering also observes that Gipps' idea that the concept of illness is rather intuitive is more like a "pretheoretical idea", or perhaps some form of "reaction" (Pickering, 2003, p.274) that is ultimately based on the idea that there is something constitutive and obvious in the notion of mental illness. As I argued, current philosophical accounts are far from considering obvious what the correct description is. Also, the very fact that there is disagreement about mental illness shows that this agreement is not as taken for granted as Gipps assumes.

In summary, Gipps' main objection, that is, that the notion of mental illness is classified in a certain category due to them being constitutive of our previous notions of health or illness, therefore being intuitive is, ultimately, ill-founded; Gipps does not acknowledge that what it is supposed to be obvious is actually an ongoing debate, and what is considered constitutive can change over time.

e) Response to Thornton and Loughlin

Tim Thornton, in his (2007) book proposed that Pickering's argument is either trivial or relevant only for undermining foundationalist accounts.

> His [Pickering's] central claim is that detectable and observable features of a condition, a putative illness, cannot be described without begging the question of the pathological status of that condition. This is not, however, a surprising claim. If the correct description of the features is taken to imply a pathological status then, trivially, it cannot be independent from the overall status. Even if it merely provides evidential support, Pickering has really only undermined a foundationalist version of the likeness argument (…) (Thornton, 2007, p.29)

Thornton, in any case, accepts Pickering's claims on epistemological grounds, although it dismisses it as trivial. Alternatively from being trivial, the other option is that, according to Thornton, foundationalist views are affected by Pickering's central claim. A foundationalist view is "one in which observational data can be established independently of any broader theoretical perspectives" (Thornton, 2007, p.29). So, Pickering's point is, at least, damaging for this kind of account. But, both Thornton and Pickering agree in the fact that the only real options are holistic views, due to the fact that foundationalism is often seen as an outdated and implausible view of science. Thornton argues that holistic views are immune to the objections against the likeness argument. This can be interpreted in this way: the actual ontology is unaffected by the epistemic theory ladenness of the features. It seems clear to me that Thornton concedes that the epistemology is theory laden given that the weight of the features is determined by the meta-theory used. But, Thornton maintains that holism would save the ontology from the question begging that Pickering says. This means that either the classification

overall can be independent of the epistemology, or that holism cancels the possible problems with the classification.

In the first option, where epistemology is independent from classification, I concede that this can be the case in some very specific occasions. This usually happens when the phenomenon is an unexplained singularity, for example, black body radiation in physics before quantum physics came about; or when the phenomenon is sui generis. Mental illness, although considered out of the norm, is not something outside of the theory that we are trying to describe. Mental health professionals consider to have a grasp on what the phenomenon is, for example, depression. Neither is it considered sui generis, because it encompasses a whole spectrum of phenomena, and not all of them can be sui generis.

The second option, which I think Thornton is more sympathetic to, is that holism cancels or minimizes the possibility of being left only with the ontology that the meta-theory is presupposing. I think that having multiple competing theories, which are question begging all the same, does not make the problem go away. I think this is what Pickering had in mind when he designed his examples about alcoholism and schizophrenia. Both start from different assumptions and end up with different epistemologies and different ontologies. A multiplicity of theories is not helping to decide which one is correct.

Thornton also argues that the theory ladenness of the epistemology can actually be independent from the proposed ontology:

> The holism that Pickering highlights only establishes an epistemological point: that we cannot establish or know the nature of the features of a condition independently of establishing or knowing an overall classification. This is not the same as saying that the features are constituted as the features they are through human judgement. (Thornton, 2007, p.30)

I address this specific worry later on in this section when I emphasise the generality of the argument and the issue of categorizing.

Summarizing, Thornton's main worry is that the likeness argument is either trivial or just affects foundationalist views. His objections are aimed to disarm Pickering as a general point, not necessarily in the context of mental illness. He thinks that Pickering's

objections against the likeness arguments are familiar to philosophy because they resemble the theory-dependence of observation.

Before addressing Thornton's worry I will consider first an objection from another author, Michael Loughlin, because his objection is related to Thornton's claim, and then I will consider objections from both authors within a more general point.

Loughlin's (2003) main concern is that, even if there is no agreement in the debate on mental illness, still many accounts have coherent, rational arguments. Loughlin says that agreement is not the only goal to achieve. In particular, Loughlin argues that the likeness argument is useful to analyse rival accounts in order to discover their underlying assumptions, and this will eventually bring a refinement of the theories, possibly up to the extent that everyone might be satisfied in the future (Loughlin seems to imply the latter, although he does not say it specifically). In this sense, he defends that the likeness argument is not a failure, but rather, a useful tool.

Against this view, I am not convinced that one should just take the pragmatic view of considering all accounts as more or less equally coherent and work with that. They are so deeply incompatible that there is no common ground in which one can argue with in the first place. In fact, my main argument about not accepting coherence as a redeeming feature is that in my view they are all equally wrong (or equally right). There is not a clear winner. Loughlin's whole paper, however, seems to be dedicated to discredit the idea that disagreement is a problem. To him, disagreement is a natural consequence of things and will be naturally dissolved by the use of the likeness argument itself.

In response to Loughlin, the matter is that he concedes the logic of Pickering's criticism; i.e. he agrees that likeness arguments are circular, but then contests that this makes them useless. This latter point is advertised when he says "Yet the precise force of such considerations is unclear, especially if they are being used to support the view that observations cannot even "help decide" the issue" (Loughlin, 2003, p.262). So, Loughlin thinks that, despite their circularity, there is some purpose in likeness arguments. That is, that the fact that they are circular doesn't have the force Pickering thinks it does to suggest that they are useless. He says:

> So of course the likeness argument cannot bring closure to the debate, but (those criticized for using it might complain) it is quite unreasonable for Pickering

> to expect it to do so. Pickering proceeds as though an argument has failed if it has failed to produce agreement among all parties to the debate. But this notion of what it is for an argument to work is untenable. (Loughlin, 2003, p.263)

That Loughlin thinks that the likeness argument is circular, and therefore useless, is perhaps implicit in the way he thinks that Pickering's is a familiar point (in a similar way that Thornton thinks so as well):

> Pickering relies in part upon a point that has been well established by extensive debates within the philosophy of science. Simplistic forms of deductivism are false, because there must always be a role for judgment in the interpretation of scientific evidence. (Loughlin, 2003,p.263)

So then the aim of Loughlin's paper is to dispute that Pickering's criticisms undermine the rational usefulness of likeness arguments.

In the first place, one element of Loughlin's answer is to distinguish between the idea that an argument does persuade and that it should persuade; he calls this a 'moral' point. But this is not entirely valid, because in the context of mental illness, it is not clear that a likeness argument should rationally persuade if its premises beg the question against a prima facie equally plausible view.

Summarizing: Loughlin argues that the likeness argument is circular, but that does not need to be a disadvantage at all because disagreement helps philosophical thinking by questioning the background assumptions of the different theories to assess the account's viability.

Pickering's defence from Loughlin is focused on the issue that disarming the likeness argument does not imply that questioning the relevant philosophical views is not possible anymore. Pickering's emphasis is that the facts do not cease to be the main force, he explains that only the form of the argument that the debate has taken that is problematic:

> [T]he interpretations I speak of are supposed to be of the scientific facts about alcoholism and schizophrenia, for example, of the accepted facts of brain chemistry and so on (...) These facts constitute important insights into the nature of the brain and its workings, albeit they may be based on a historically limited

> set of conceptual resources. It is what we make of these insights that is at stake, I believe, in the debate about the reality and nature of mental illness. (Pickering, 2003a, p.275)

Along this section I have considered some of Thornton's and Loughlin's specific points against the likeness argument; however, I still have not argued against the core of their arguments, which they share. Loughlin and Thornton make the same point about the logic of the argument: both contend that Pickering's objections *are a claim of circularity*, that can trivially be blocked in Thornton's case, or that is simply not damaging at all in Loughlin's view. This reflects a deeper problem that might look as Pickering's objections are definitively unviable.

The deeper worry comes from the following potential objection: if Pickering's criticism of the likeness argument is successful, given that the argument seems like a very general argument of how classification works, then it threatens to make classification in general impossible. An instantiation of this worry could be like this: if one assumes that the objections to the likeness argument are valid for all sorts of classification, and therefore the likeness argument is invalid per se, then even trivial examples of classification are deemed invalid because of its circularity (as Thornton and Loughlin would claim).

My response to this potential objection is that, although it is the case that Pickering suggests that his diagnosis is about the form of the argument, something that he does not emphasize, but is playing a role, is that there is live opposition (that is, there are contending options) in the case of the reality of mental illness. For example, someone could say that alcoholism *is like* an illness because it has symptom-like phenomena, whilst someone else could say *it is like* a moral weakness because it has features of moral failure *in the context of that actively contested debate*. In other words, it is only due to there being a live debate that the likeness argument and its objections are relevant.

By contrast, in cases where there is agreement by the contesting part, the likeness argument becomes irrelevant. If somebody says that something with an atomic number of eight is classified as oxygen, the likeness argument is being used, but since there is no live opposition, it makes no sense to bring the objections to the likeness argument to the discussion.

Coming back to Thornton and Loughlin, my point about their core objections against Pickering is that they missed the point because Pickering's is not just a general claim about the theory-dependence of observation or holism. Pickering's point is that his objections can be applied to a particular debate where there are principled reasons for thinking that there are resources to maintain different sides. Loughlin and Thornton can both be undermined by saying that, whilst likeness arguments can be helpful in other areas, Pickering has highlighted a flaw *in the way they are appealed to in this debate* because the circularity in the descriptions of 'symptoms' and then, inferring illness, is too demanding.

More specifically, Thornton's worry is that Pickering's point is trivial and that holism will dissipate the objections to the likeness argument because there is no such thing as raw data. Theory will affect data, and vice versa: data will affect theory. And Loughlin argues that the likeness argument is circular but still useful as a philosophical tool. But both are missing the point because Pickering is not merely making a logical claim. Rather, it is an illustrative point that has to do with the description of the features. Both objections are blocked by the fact that this is a context specific debate, where all the possible resources are open ended.

However, it is not clear whether Pickering himself intended to make merely a logical point against the likeness argument, or whether Pickering himself realised that his objections are only relevant to claim contingent to the specific ongoing debate. However, in my view, Pickering's objections are valid due to the fact that in the debate of mental illness there are currently live possibilities for alternative conceptualisations of the discussed conditions. Pickering offers a diagnosis of why using the likeness argument will not resolve a live debate. But had there not been a live debate – as in the case of oxygen and atomic number - the diagnosis would be irrelevant.

Summarizing, both Thornton's and Loughlin's criticisms fail because, as Pickering intended it, his objection operates only in the context of an actively contested debate. In other debates, Pickering's claims do not apply because in the vast majority of conceptualizations there is not live opposition.

III. Clarification on how to interpret Pickering's view

At first, the reality of the mental illness debate looks like many other debates in philosophy where there is disagreement between the parties. However, I argue that the debate about the reality of mental illness is different. It is fundamental for Pickering's argumentation to work to note that his claim is not a general point about the circularity and theory ladenness of philosophy in general; it is contingent because of the following reasons.

The question here is: How to settle the debate? Due to the form in which Pickering's argument works, the point is that, given a certain criterion, philosophers have to decide which one is the most important or more essential feature in the concept of pathology. Then, if any of such criteria were the most essential according to all the debating parties, the debate would be settled in favour of the account that is closer to those criteria. However, as Pickering correctly argues, the fact that their preferred theory influences their choice of the criterion itself shows circularity in each author's reasoning about the reality of mental illness. But, the important issue to note here is that the problem does not consist in the circularity itself, as it might seem from a first glance. The real problem, specific to this debate, is that there is not an objective way that is neutral to the dispute to decide between the different existing criteria. Therefore, Pickering's conclusion follows: It is inconclusive to use the likeness argument to generate agreement in the reality of mental illness debate.

Nevertheless, it is worth mentioning that Pickering's inconclusiveness of the likeness argument is limited to the context of the debate. It is possible that in other debates similar arguments from analogy may actually lead to agreement. That can be the case when there is agreement from the parties about the relevance of the essential features of a phenomenon. It follows that, when comparing two competing accounts, one can determine which of the accounts is closer to the given guideline on the basis of the likeness of the features of each account. Such a strategy might succeed in convincing all parties about the nature of the phenomenon, as the conclusion of the argument from analogy is valid for this case. Therefore, Pickering's diagnosis about the debate is meant to be applied specifically for the context of the debate of mental illness.

It is unclear whether Pickering designed his claim in order to be generalized to all the disputes in philosophy that work with a similar argument to classify a certain phenomenon. The reason of why there is doubt about the reach of his account is that, although he uses his claims only within the context of mental illness, he never specifies the reach of his rejection of the likeness argument. It could be the case that he meant this to be a general criticism towards philosophy and its way to classify through the likeness argument. However, it is clear that his account would be too hard to generalize to entirely different debates: there is the intuition that either disagreement is not a problem in some philosophical cases (for example, internalism versus externalism), or the likeness argument becomes too trivial in other cases (for example, in classifying everyday objects). However, in the context of mental illness, and in the case of delusions too, Pickering's claims make an adequate diagnosis for the current disagreement.

Summarizing: Although he did not make it sufficiently explicit, Pickering should have said that his claims are contingent; that they can only be applied to the mental-illness debate. The flaw in the likeness argument is not just about the circularity in itself, rather, the flaw is that the circularity starts with some premises that are exactly what is contested in the specific debate.

So, what I have done now, is that I have extracted a principled reason of why in a specific area of philosophy there is ongoing debate about the reality of mental illness, but that the mechanism that philosophers use to try to reach an agreement in that debate, that is, the likeness argument, cannot work. Therefore, in the second part of this chapter I am going to argue that the same applies to the case of delusion.

IV. The likeness argument in delusion

In the previous sections of this chapter I have looked at Neil Pickering's account of the likeness argument, some of his critics, and how he should have responded to his critics. In the next sections I am going to illustrate the way the resultant analysis of the likeness argument applies to delusion. In other words, through these sections I explain how is it that the likeness argument sheds light on the ongoing disagreements about the nature of delusion. One of the first considerations in applying the likeness argument to this case

is that Pickering's central question is not exactly analogous to the case of delusion. The question here is not whether delusions exist or not, or whether or not they are symptoms of mental illnesses. The question is: What is the most adequate characterization of delusion? The form of the question is important because it has consequences for the scope of the possible answers. If the question were only whether delusion is a mental illness or not the possible answers fall into two possible ways of categorizing it: It is a mental illness, or it is not. The justification for any answer, as Pickering shows, is complex, but there are still only two possible outcomes.

In the case of delusions, however, there are several different possible outcomes. Consider the following: so far, I have classified different accounts of delusions according to their main assumptions regarding the question: 'what is the true characterization of delusions?' Doxastic accounts hold that the answer has to do with what sort of beliefs delusions are. In more general terms, doxastic accounts argue that delusions are beliefs whilst anti-doxastic approaches deny this. On the other hand, phenomenological positions examine, in general, what sort of (phenomenological) experience constitutes a delusional state. The contrast between these two positions shows that the problem of characterizing a delusional state leads to an undesirable ambiguity (or underdetermination): when analysed through the structure of the likeness argument, different accounts propose different ontologies. In the general case of mental illnesses, the likeness argument is more specific because it is applied to a narrower case than this; however, I consider that the form of the argument – along with the two main assumptions supporting its two proposed objections – is essentially the same.

In this part of the thesis, Pickering's defence against Tyreman's objection (discussed in the previous section), becomes relevant. Tyreman (2003) defends his point in the debate about mental illness by selecting Fulford's (1989) account as the correct one. But Pickering's point, with which I agree, is that *even if there is a correct account*, it still leaves open the question for any condition whether it is question-begging regarding its description. So, the way the symptoms of the phenomena in question are characterized would still be an open debate. In other words, although Pickering is advancing a logical argument about the shape of the likeness argument, there is an additional feature that is important in the debate about mental illness, which is that it is actively contested and, there are to be resources on the opposing sides of the debate to frame the description

of the symptoms in equally consistent, but opposing ways. So, it is Pickering's logical point, but also the specific contest of how the debate works, that actually makes his objections to the likeness argument relevant both for mental illness (in the previous section) but also for delusions (in this section).

Within the previous chapters of the thesis I have shown that there are various potentially plausible accounts of delusion, but also, that they are all subject to strong counterarguments. Therefore, it is clear that there is widespread disagreement about the kind of states that delusions are. I am going to explain with some instances of delusion, taken from the previous chapters, why I think there is an ongoing debate in these cases. My point is that rival accounts in delusion take the same form of argument that Pickering has highlighted; these rival accounts are instances of the likeness argument.

As I set out earlier, given that the form of the likeness argument is taken by accounts that judge whether a phenomenon is sufficiently like other phenomena whose categorization is paradigmatic, then, it can be said that the likeness argument, in its generic form, works by assessing whether a certain phenomenon is like some other by comparing the features to the ones of the undisputed phenomenon. If the features are enough like the mentioned canonical phenomenon, then the phenomenon in question is deemed similar enough to belong to the same category.

Recapitulating, in order to show some sort of objectivity, the likeness argument assumes three things in its generic form; those are the two assumptions that Pickering sets out for the mental illness debate, but also, that the debate must be one in which both sides have competing plausible characterizations. These are, in its generic form:

1) That the features of the phenomenon decide what category, or kind, these conditions are a member of.

2) That, with respect to the presence or absence of such features, a certain phenomenon is describable independently of the category it is assigned to.

3) The debate has to be contested, and that there have to be resources for coherent but rival characterisations of component aspects.

Applying this abstract form to delusions the likeness argument looks like this: Delusions will rightly be classified or categorized as similar to a certain positive account to the extent that they have the features of the proposed ontology; and to the extent that they lack these features, classifying them so will be a mistake, given that there is an ongoing debate about its nature.

My suggestion is then that all the accounts analysed in previous chapters take the form of the likeness argument. In sum: Currie (2000) maintains that delusions are sufficiently like imaginings. Graham and Stephens (2007) propose that delusion is sufficiently like a second-order mental state. Bortolotti (2010), Reimer (2010) and, Bayne and Pacherie (2005) argue that delusions are sufficiently like beliefs. Gallagher (2004a, 2009) proposes that delusions are sufficiently like alternative realities, whilst Ratcliffe (2010) maintains that delusion is sufficiently like a personal, non-shared, subjective experience governed by the rules of the horizon of possibilities. The positive cases proposed by Sass (1994), Berrios (1991), and Campbell (2001) have the same characterization: Sass tries to show that delusions are sufficiently like solipsism, Berrios that they are sufficiently like empty speech acts, and Campbell has said that delusions are sufficiently like Wittgenstein's framework propositions. I will have further specific examples throughout this chapter.

V. Instances of the likeness argument in the case of delusions

As I argued above, the likeness argument form applies to all cases in my thesis. It would be redundant to go through them all, so I will select one or two from the different categories that I have highlighted in the previous chapters.

Antidoxastic accounts

Antidoxastic accounts propose that delusions are not instances of belief or anomalous beliefs: they are other propositional attitudes. Rewritten in the form of the likeness argument, delusions are instances of non-doxastic propositional attitudes because they are sufficiently like imaginings, second-order thoughts, bimaginings, or other propositional attitudes; further, delusions are not sufficiently like paradigmatic beliefs.

Within the framework of the likeness argument form, delusions have the following characteristics in antidoxastic accounts:

1) Delusions are better described as propositional attitudes that are not belief-like –Anti-doxastic accounts say that the odd characteristics of delusions are better explained by viewing them as instances of other propositional attitudes rather than beliefs.

2) Antidoxastic positions argue that the concept of belief cannot be fitted to delusions but then typically tailor a different propositional attitude to fit the features of delusions.

3) Delusions are not like beliefs because they are inert in many behavioural aspects, rationality in delusions is atypical, and evidence plays a limited role in the formation and maintenance of delusions- These characteristics are crucial to non-doxastic descriptions and play the role of evidence against doxasticism and in favour of the antidoxastic description.

4) Delusions are propositional attitudes rather than empty speech acts, for example, because the subject tries to express some content through their utterances – Delusions are imaginings, or second order propositional attitudes, or a mixture of the functional roles of two known entities (an imagining and a belief) because their features are better described are like instances of its subsidiary properties, including its content.

In this case one can see that the main features of delusion as proposed by Currie, for example, may be described in the likeness argument form: although imaginings are capable of having many of the inferential characteristics of beliefs they are more easily triggered by perception than beliefs; imaginings can have wild hypotheses in response to an odd experience, and imaginings are not apt to be revised in the light of evidence. Therefore, for Currie, delusions are imaginings because they have sufficient features of imaginings, and delusions are not beliefs because they lack many of the necessary features for belonging to that category.

However, there is one proposed antidoxastic analysis of delusions, discussed earlier in this thesis, which may not seem to fit the form of a likeness argument. That is Egan's (2008) claim that delusions are 'bimaginings'. Since the class of bimagining has been designed specifically to fit delusions it may seem that this is not an application of a likeness at all. However, bimaginings are a combination of functional roles of paradigmatic beliefs and paradigmatic imaginings. Therefore, in this case, the likeness

argument takes the following form: Delusions are like a mixture of functional roles of two different propositional attitudes, which are intelligible on their own. In other words, delusions are sufficiently like this new category because they fit these pre-existent functional roles. Even if it is not straightforward, Egan's account also falls under the general form of the likeness argument because he suggests that bimaginings are (or, are like) some sort of propositional attitude that does not fit a conventional 'boxological' approach to standard forms of attitude.

Regarding Egan's position in the ongoing debate, he suggests that delusions are a mixture of functional roles from beliefs and imaginings. However, given that Egan is proposing a *new* propositional attitude, different from paradigmatic belief, he is plainly antidoxastic; his role in the dispute is as one of the available options against the belief stance.

Doxastic accounts

Doxastic accounts suggest that delusions are beliefs because their features are sufficiently like those of beliefs. Although delusions might be more irrational than everyday beliefs, or perhaps anomalous, the important features of the belief status, those that count more towards their classification, remain. These accounts agree in the following aspects:

1) Delusions are better described as a belief rather than some other propositional attitude – Doxastic accounts say that, although delusions do not have the features of paradigmatic beliefs, the features that delusions present are sufficiently belief-like to matter for their classification as beliefs.

2) The belief concept is flexible- Instances of everyday beliefs, and delusions, share common features.

3) Delusions are beliefs because they affect behaviour, speech, and emotion in many cases.

4) Rationality failures in delusions are not different in kind from other beliefs, like in nihilistic beliefs or in reasoning biases.

5) The content of the delusion is important- Patients endorse and understand what the delusion's content is. They might realise it is odd or implausible to other people. Patients sometimes develop other beliefs to support their delusion too.

As an example of doxastic accounts, I will use Bortolotti's view of delusions as irrational beliefs. In her account, delusion is no different in kind to other beliefs because the same sort of rationality portrayed in patients also fails in the case of normal beliefs. To her, for every instance of rationality (procedural, epistemic, agential), it can be shown that normal beliefs might be classified as irrational by the same terms. Bortolotti's conclusion is that delusions lie in the irrational part of the spectrum of beliefs. This implies that everyday beliefs must have the features of delusions; Bortolotti argues that this is the case because normal beliefs have irrational features too. Therefore delusions are like other everyday beliefs.

Phenomenological accounts

Phenomenological accounts aim for a different description of the nature of delusion when compared to analytic traditions. Delusions are an experience because the important features are the ones concerned with the patients' subjective experiences. Phenomenological accounts share the following features:

a) The concept of *belief* is not central to the nature of delusions- Phenomenologists do not deny the possibility of delusions involving beliefs, or another propositional attitude.

b) The importance of subjective experience- Phenomenologists describe the experience of the subject through concepts such as delusional atmosphere, horizon of possibilities and alternative realities where subjective and personal features are present.

c) Understandability- Contrary to Jaspers, contemporary phenomenologists argue that delusions are understandable through a more encompassing concept of empathy.

d) Empirically adequate explanations- Phenomenologists such as Gallagher and Rattcliffe maintain that their views are complemented by empirical explanations of the features of the phenomenon.

I will exemplify the likeness argument here with Ratcliffe's account. He maintains that delusions are a subjective experience which can be understood through the encompassing concept of radical empathy. The features that matter for understanding

and categorizing delusions are the subjective experience, the patient's affect, and the capacity of simulating another person's experience (that is, radical empathy). These latter, and familiar phenomenological characteristics, are the ones that count towards the classification of delusion as a phenomenological experience. The features that describe delusion as doxastic, or not, are unimportant such as the delusion's connections to rationality or behaviour. In other words, a delusion could be a belief or not, but that it is not important for its description as delusion.

Ratcliffe is widely discussed amongst phenomenologists. His account, and the ones from other phenomenologists, are not usually discussed in analytic circles because they do not share common ground. And, as many phenomenologists do, he advances his view against Jaspers due to him still being highly influential. Therefore, delusions are a particular kind of phenomenological experience because they can be explained through phenomenologically characterizable features like delusional atmosphere, empathy and understandability.

Meaning accounts

Sass, Campbell and Berrios address an important issue: they ask whether delusions have content by analysing their meaning. But, as I mentioned in the conclusions of Chapter V, their search for meaning comes to an unexpected halt due to their own assumptions about rationality. The three authors I analysed in this chapter agree on the following features:

a) Meaning will define the delusion's status, its nature: a deviant framework proposition, a solipsistic declaration, or an empty speech act.

b) Meaning in delusions is not straightforward; in other words, what patients say is not exactly what they mean. The delusion's meaning is either rather strange, deviant, or null.

c) Rationality defines meaning- For Berrios, since there is no rationality in delusions, they do not have meaning. For Campbell the utterance does not have meaning either, although it has some kind of sense as a framework proposition. Sass argues that the patient is broadly rational, but within a special notion of rationality, like that observed in solipsism.

These features decide what delusions are: deviant framework propositions (Campbell), a solipsistic expression (Sass), or an empty speech act (Berrios).

In this case I will use Campbell's account of delusions as framework propositions. For him, delusions might or might not be beliefs, but the important issue is whether one can ascribe meaning to them. In the case of the Capgras delusion, one cannot say what the patient means because, in whatever way one can interpret the assertion, the meaning-ascription fails. This failure is due to the fact that the patient does not engage in canonical ways of testing his thoughts. This implies, for Campbell, that the meaning is lost because of an incompatibility with the Davidsonian constitutive principle of rationality. Campbell's conclusion is that particular instances of delusion are not rationally interpretable. However features such as not being subject to testing would make sense if they were understood as akin to framework propositions. Hence the idea that delusions are deviant framework propositions because they share sufficient features of ordinary framework propositions.

Having reviewed that the likeness argument applies to instances of delusion, I will exemplify how the weak and strong objection to this form of argument apply too.

VI. Objections to the likeness argument in the case of delusion

If one accepts that the previous accounts of delusions have the form of the likeness argument, then that implies that they also rely on the two assumptions discussed in the previous section:

1) That the features of delusion decide its categorization – For example, features such as the patient's lack of action count towards categorizing it as an imagining. Another feature, such as the patient's conviction is decisive to categorize delusion as a doxastic state.

2) That, with respect to the presence or absence of such features, delusion is describable independently of the philosophical account it is assigned to – Delusion, ideally, has to be describable independently whether it is classified as doxastic, phenomenological

state, etc. For example, definitions of delusion in the DSM try to be independent from the philosophical framework.

3) The debate has to be contested, and there have to be resources for coherent but rival characterisations of component aspects.

I argue in the following subsections that the form of the various arguments discussed in the case of delusion are subject to the weak and strong objections.

a) The weak objection in the case of delusion

The weak* objection suggests that, in order to judge the relevant phenomenon as belonging to a certain category, people evaluate the features as having a particular weight according to their own prior theory. I will argue that, given the same framework, that is, having the same assumptions, philosophers assign different weight to the relevant features, and therefore, get contrasting results. A paradigmatic case where this is clearly visible is the doxastic versus antidoxastic debate. This is a particular case in which people agree in most of all the fundamental framework assumptions. However, both accounts end up with opposite outcomes.

Firstly, I will summarize the doxastic and antidoxastic shared assumptions: Both agree on the fact that the fundamental question is whether delusions are, or are not, beliefs. They both take functional roles to be decisive to determine the belief-status of the delusion. Neither is concerned with questions about the meaning of the utterances to which the delusion gives rise, nor its status as an experience. They both allude to functionalist examples and, contrary to phenomenological accounts. They propose that the features should agree with the vision of a shared, objective, world. Both accounts aim to be empirically adequate and to "explain the phenomena". Both use a flexible concept of rationality, where functionalism is privileged over other explanations. All these shared assumptions show that both the doxastic and the antidoxastic accounts share the same broad background framework.

Secondly, despite this background shared agreement, they differ with regard to two fundamental things: 1) how they weight the evidence and 2) what constitutes a belief (i.e. are beliefs rational? are beliefs acceptances?). I will explain this further:

About 1) I will analyse an instance in which both they both disagree about the same feature. In antidoxastic accounts, the lack of action of the patient counts against the concept of delusions as beliefs. In Currie's account, delusions are more like imaginings because in the latter no action is required. In Egan's account, delusions have some of the characteristics of non-beliefs, and the lack of action is one of them. In Graham and Stephens, the lack of action is also an indication that delusions are second-order thoughts, because the latter usually do not lead to action. In doxastic accounts, the lack of action serves the opposite purpose. To Bortolotti, the fact that patients do not act in accordance with their delusions is a token of how this is similar to hypocritical normal beliefs. To Reimer this is an instance of how delusions are similar to nihilistic philosophical beliefs. Therefore, the lack of action is a feature that counts as evidence for two opposing purposes that share the same framework over all.

About 2), it is useful to observe that the concepts of belief used in delusion accounts are different, especially between the doxastic and antidoxastic accounts. The concept of belief varies from having epistemic, procedural and agential rationality to having functional roles. The antidoxastic account of delusion varies from being an imagining to being a second order thought or delusional stance. These count as antidoxastic because it is assumed that the concept of belief could not fit these roles. This possibility of rational disagreement about the concept of belief is what makes possible the difference in assigning weight to the characteristics mentioned. Were the characteristics of the concept unified and unambiguous such difference in weighting the evidence would not be possible.

Therefore, one can conclude that in the case of antidoxastic and doxastic debate although both accounts of delusion share the same broad framework, that is, they have similar shared assumptions, they still disagree on the nature of the phenomenon because of the conceptual theory they use to analyse it. This shows that their disagreement is mainly due to being liable to the weak* objection.

How does the weak objection work in delusions? To be specific, I will use the accounts form the previous section to illustrate it.

a) In Currie's account delusions are imaginings; Currie considers the following characteristics as crucial:

- Imaginings are capable of having many of the inferential characteristics of beliefs.
- Imaginings are more easily triggered by perception than beliefs.
- Imaginings can have wild hypotheses in response to an odd experience.
- Imaginings are not apt to be revised in the light of evidence.

In this view, delusions are imaginings rather than beliefs because the weight given to any of those characteristics, for example, "they are not apt to be revised in the light of evidence" are considered more important than the other characteristics like "they connect with behaviour" or "they affect other beliefs". Therefore, Currie's account is vulnerable to the weak objection because there are features that matter more for the characterization rather than the ones available in other theories.

b) The same argument applies to consider how the weak objection works in Bortolotti's account; for her, the belief-like behaviours count more towards classification of delusion than the features that are not like paradigmatic beliefs. Even more, her account goes even further by modifying the notion of rationality in beliefs to accommodate her delusions' view. In her case it is not only the delusions' features what counts, but rather, her view also gives more weight to the irrational part of everyday beliefs.

c) Ratcliffe's account is more radical regarding the weak objection because his view disregards most of the features that analytic accounts consider important for classification. Ratcliffe, (as most of the phenomenologists) assigns weight to features of delusion that have to do with delusional atmosphere, empathy, and the patients' subjective experience. He disregards the content of delusion, or whether it is rational or doxastic. His view considers that delusion is an understandable experience due to the phenomenological features it portrays.

d) For Campbell, the main feature of delusion is that its meaning is puzzling; it cannot be adapted straightforwardly to his notion of rationality. Therefore, Campbell suggests that the puzzling meaning of delusion can only be understood through it being held immune to testing, in the same way framework propositions are. In this way Campbell considers the following features important for classification: rationality, meaning and

content. Campbell thinks that by solving the puzzle about how to make sense of these aspects of delusions he will be able to describe them.

The discussion above shows how particular features of delusion can be weighted in different ways in order to account for their own theory, therefore, ratifying the weak objection.

b) The strong objection in the case of delusion

The strong objection* against the likeness argument is the following: the particular framework will influence the features or characteristics assigned to the phenomenon, therefore begging the question from the beginning. This broad characterisation can be better exemplified by taking particular accounts. There are two possible instances of the strong objection: How a particular framework leads to question begging, or, how different assumptions may lead to incommensurable results between accounts, even if they are about the same phenomenon. In this case I apply this objection to the particular cases I discussed above.

a) Given Currie's explanation that in delusion there is a defective part of the brain (whose function is to monitor the efferent copy of the patient's movements), it could follow that such a mechanism might affect the belief system, or any other system of the brain. That is, because brain damage can virtually affect any function of the brain, the option of having damage that affects belief, even if indirectly, is an available explanation. However, Currie justifies his account in a different way, that is, by proposing that an imagining is more adequate than the belief definition of delusion. This seems unfounded. Although Currie had different options available, he chose to describe delusions as imaginings and to endorse a full antidoxastic stance. This is due to the prior assumptions he made about the weight of the features (due to being influenced by the weak objection); he already had an idea in mind that begged the question and therefore it influenced the question he wanted to answer.

b) The main question in Bortolotti's account can be summarized as "Is delusion sufficiently different in its rationality to other normal beliefs in order to count as a different natural kind?" Bortolotti provides a twofold answer. The first step consists in showing how normal (although, non-canonical) beliefs are irrational. The fact that she uses non-canonical beliefs as examples of irrationality becomes important in the analysis

of incommensurability. The second step in Bortolotti's strategy is to show how delusions are not completely irrational given that they adhere to the same logic as normal people, fall under the same normal biases, and show equivalent behaviour to that of normal control groups. Now, the question is: How is this framework theory laden?

First, Bortolotti's notion of rationality is a flexible account that is determined by accommodating forms of thinking susceptible to having failures. For example, her form of agential rationality is such that some instances of normal, canonical, human actions are irrational, as is the case in forms of simple hypocrisy. This is the base of Bortolotti's notion of rationality as a continuum[22]. Therefore, it is clear that the notion of continuum is already in her defining criterion.

Second, Bortolotti defends the rational part of delusion. (The latter could have been otherwise, and hence, antidoxastic accounts are born). The fact is that, given her account of rationality, her only choice was to be doxastic or antidoxastic. And, as we have seen from the weak objection, to be the former or the latter is only a question of how much weight one assigns to features of the instance one needs to classify. In this case, then, it follows that both the weak and the strong argument work together to show the theory ladenness of the application of the likeness argument, and that results in Bortolotti's defence of the rational part of delusion.

c) Ratcliffe's main claim, that delusions are understandable through radical empathy, is trying to maintain a balance between the bizarreness of delusion and the fact that it seems to make sense in some aspects. His answer is a reaction against Jaspers' definition of delusion as ununderstandable. Ratcliffe's question is embedded within this framework: he disagrees with the conclusion drawn from Jaspers, although both use the same concepts to describe delusion. This account is a good example of how the strong objection works because one might erroneously think that given a specific framework, the result is fixed. But Ratcliffe's account is an example that, even within same framework accounts, stances may disagree. The question begging here means a different thing; it means that the question, 'what are delusions?' is preceded by a theory

[22] Note that I am not implying that Bortolotti's account is not objective. The latter assertion is supported by the fact that Bortolotti might have decided that delusions were not able to be accommodated in her notion of continuum, and therefore, she could well have been an antidoxastic, although due to her primary question, always under the rules of the doxastic vs antidoxastic debate.

that gives a specific form to that question; in this case, are delusions an understandable experience? The answer, however, can have many different forms (as many as delusion accounts exist), but the question is already setting the answer's parameters. Ratcliffe's account is that delusions are indeed understandable; Jaspers' is that they are not. However, an account of, say, doxasticism, would have to use a different theory and different parameters and would get to a different conclusion.

d) Campbell's fundamental question is: Are delusions understandable under the constitutive principle of rationality? To account for that, it is necessary to assess whether the subject engages in canonical forms of checking his assertions. It comes from empirical findings that patients do not do that. Therefore, under his own terms, it has to be the case that any particular instance of delusion is not understandable. It is Campbell's notion of rationality that dictates a priori whether something is understandable or not. In Campbell's account, the notion of rationality is less flexible, and therefore, no other choice arises from his framework. In that sense, his question of meaning is analysed under his parameters of rationality and so his answer is consistent with those assumptions.

Summarizing, with these examples I have shown two things: that the premises of each question determine the sort of account obtained; and that by the light of each account, even when looking at the same phenomenon, with the same feature in mind, different frameworks lead to different results. This suggests that the strong objection to the likeness argument applies.

The fact is that, although clinicians have tried to formulate their delusion account in an atheoretical way, classification necessarily has a theory behind; given the fact that the nature of delusion is an ongoing debate, theory will lead to incommensurable results. There is no possible way in which an account is objectively better than other, or that there is an uncontroversial winner given these conditions. I analyse this in the following section.

c) The live debate

Summarizing what I said earlier in the chapter, an important part for Pickering's objections to the likeness argument to work is that there must be an ongoing debate with the possibility of rival but internally consistent descriptions of subsidiary features.

The real problem, specific to delusions, is that there is not an objective way that is neutral to the dispute to decide between the different existing criteria. Therefore, it is inconclusive to use the likeness argument to generate agreement in the reality of mental illness debate. One might argue that the objections to the likeness argument are too general. However, this is not the case, precisely because disagreement can only be possible in a context where both sides have principled arguments to maintain an ongoing debate, as I argued earlier on about Pickering's account.

The fact that classifying delusions is an ongoing debate is clear from the previous chapters. Not only does each account have its detractors and criticisms; also more general frameworks (for example, antidoxasticism) have been contested. I will again draw an example from each chapter to show how this works in particular accounts of delusions. In the case of the debate amongst general frameworks, although there is explicit discussion amongst them (some of it is mentioned in my thesis, for example, Ratcliffe against doxasticism, or Bayne and Pacherie against antidoxasticism), each case is different and I think that trying to confront one aspect of the debate against the other would end up being too general. Therefore, I will only focus on particular accounts.

a) In the case of Currie, the ongoing debate is against the description of delusions as beliefs. Currie argues that delusions are a particular form of failure in the brain's monitoring functions, and that it is impossible to empathize with them. Currie suggests that the view that delusions are beliefs could be replaced by his account of delusions being imaginings that the patient mistakes for beliefs. His main detractors come from doxastic accounts such as (Bortolotti, 2010), and (Bayne, 2005). However, many of Currie's assumptions are endorsed by Graham and Stephens, who took one of his ideas to create their own antidoxastic account of delusions being a second order state.

A particular instance of this debate in the form of the likeness argument would be: Currie maintains that delusions are (like) imaginings because they can be explained by a failure of monitoring in the brain. Bayne argues that delusions are (like) beliefs because brain damage is not the way beliefs are formed; beliefs are formed through other beliefs or through perceptual experience only.

b) Bortolotti argues explicitly against many of the antidoxastic accounts, including (Currie, 2000), (Graham and Stephens, 2004), and Frankish (2009); however, she also

considers objections against meaning accounts like (Sass, 1994), (Campbell, 2001), and (Berrios, 1991). Many disagree with some particularities of her stance but still concede that delusions could have many belief-like features, such as (Hohwy and Rajan, 2011), or (Frankish, 2012), as I described in the section about objections to Bortolotti. In the form of the likeness argument, this can be seen as: Bortolotti affirms that delusions and everyday beliefs belong to a continuum where the irrationality varies in magnitude but not in kind. So, delusions are like beliefs, only more irrational. Frankish argues that delusions are not quite like beliefs because Bortolotti's folk-psychological definition of belief cannot capture its diverse nature. Frankish defends the view that delusions are more like a policy within his two-level framework.

c) Ratcliffe, coming from a phenomenological background, argues against empiricist accounts such as (Stone and Young, 1997), (Davies, 2001), and against Campbell's (2001) rationalistic account. He suggests that analytic parameters, such as the relation between perceptual content and belief, are mistaken. Ratcliffe disagrees explicitly with Berrios (1991) and Currie (2001) as well because, according to him, the relation between perception and belief is still uncertain. However, Ratcliffe's main opponent is Jaspers, because the view that delusions are ununderstandable is a starting point for many phenomenologists. These objections can be found in the section where I discuss Ratcliffe's view about doxasticism. In the likeness argument form, Ratcliffe claims that delusions are like an understandable phenomenological experience that can be made sense of through radical empathy. Jaspers defined delusions like an ununderstandable experience that cannot be reached through empathy.

d) Campbell, in his (2001) paper, before advancing his positive view of delusions as framework propositions, initially targets his arguments against (Ellis and Young, 1999). His intention is to make clear that delusions are not a rational response to highly unusual experiences, that is, Ellis and Young's empiricist view. He also dismisses Berrios' (1991) view because patients seem to understand what they are saying and they assert it sincerely. Bayne and Pacherie (2004) share Campbell's notion that meaning is important for classification, however, they disagree mostly with his top-down approach. Furthermore, Thornton (2008) argues that Campbell's positive view (that is, that framework propositions may shed light on delusion) is unattainable due to the role these propositions actually have: they are not an expression of any particular fact in the world.

Campbell maintains that delusions are like framework propositions because both of them are unverifiable and other beliefs are built upon this framework. Thornton argues that delusions are not like framework propositions because the latter cannot be expressed, just shown in actions, which delusions do not do.

It is clear from these accounts that the debate about the nature of delusion is diverse in the sense that authors have developed a whole range of positions against any particular stance. Disagreement may vary from a very subtle point to extreme discrepancy. At this point, accounts are very sophisticated too by considering many objections and resolving many problems that philosophers consider relevant today. However, as I maintain through this work, there is no current agreement on what delusions are, and even the possibility of there being agreement in a near future is improbable, precisely because of the non-homogeneity that pervades this debate.

The fact that all the accounts of delusion mentioned throughout this thesis are vastly discussed is not fortuitous: they have been chosen by that criterion, amongst others. These accounts purport to cover the debate in a comprehensive way, covering many of the available positions, along with some of their most important criticisms.

VII. What insight should we have about the nature of delusions?

The main focus of this thesis has been to critically analyse the various competing attempts to shed light on the nature of delusions, and to diagnose why there is not agreement amongst them. Trying to characterize a phenomenon typically takes one of two paths. We can attempt to characterise a phenomenon as basic, fundamental and sui generis. Or we can attempt to show that it is an instance of a more general and familiar kind.

In the context of a different debate, Strawson's (1958) account of persons takes the former route. Strawson argues that persons are a basic feature of our conceptual scheme, irreducible to anything else. Although he characterises them as having both mental and physical properties essentially, he does not suggest that we can have a grasp of mental properties, at least, prior to our grasp of the concept of persons. Persons are

not instances of a more familiar general concept. Thus, Strawson's account of persons is an example of an account of a phenomenon as basic, fundamental, and sui generis.

In the case of delusions one could, equally, adopt the position that delusions are a sui generis phenomenon, something unique that is not like anything else. This approach would be similar to Jaspers' definition of delusions as "the ununderstandable" that I discussed in Chapter I. (If saying that delusions are ununderstandable is an instance of the likeness argument, on the grounds that they have the features of paradigmatic ununderstandable phenomena, it is a limit case of attempting to shed light on their nature.) However, such a position is surely unsatisfying due to the combination of delusions having many recognizable features and yet remaining prima facie puzzling.

Given that the alternative is unsatisfying, it is hard to see how the philosophical project of trying to shed light on the nature of delusions (in general) can avoid the strategy of comparing them to more familiar mental states or propositional attitudes. Hence it is understandable that most authors in the debate have resorted to using likeness arguments.

What can we learn from the discussion above? I gave some particular examples in the previous section in order to show how the likeness argument works. If one assumes that a similar argument strategy is adopted in all the accounts described, there are some conclusions that follow:

i. Accounts of the nature of delusion that are not sui generis can be seen as instances of the likeness argument (for example, delusions are like bimaginings).

ii. Instances of the likeness argument form are liable to the weak objection.

iii. Instances of the likeness argument form are liable to the strong objection.

iv. The discussion is a live, ongoing, debate.

From which it follows that

iv. Other theories of delusion that have not been considered here are also likely to share the likeness argument form and will thus be liable to the weak and strong objection if there is live debate about the main assumptions.

The latter comes from a generalization of how the weak and the strong objection work: whenever a single account has more than one important feature, even within the same theory, different weights can be assigned to the features, so that some similarities can be chosen to have more importance to classification rather than others, therefore being liable to the weak objection. Further, whenever there is a question of how to construct an ontological theory of delusion, the prior question is of the sort "What does the concept need to do?", it is not the case that there are atheoretical answers to such a question, and the very same question implies some influence on its later classification, given the case that there is live debate about the phenomenon.

I have highlighted throughout this work that there are different assumptions guiding philosophical stances. My diagnosis is that there are two potential implications: the first one is that there might be ambiguity in the concepts that constitute delusion. The second issue I consider is whether relativism or scepticism are the only viable options to conclude from the patent disagreement in this specific debate.

VIII. Final conclusion- Disagreement about the nature of delusions

I would like to highlight the fact that, in accounts of the nature of delusion, diverse definitions are used to characterize related phenomena. Some examples are the following: different approaches to meaning are adopted by Sass (1994), Campbell (1999), Thornton (2008), and Read (2003). In such accounts not only is the view of meaning different; it is contradictory and incommensurable. A similar case can be made with the phenomenological concept of experience, which is remarkably ambiguous, as I explained in Chapter IV. The disagreement about the concept of belief between antidoxastic accounts is visible against doxastic accounts; even more, it is also the case that there is disagreement about the concept even within the antidoxastic group. About the ambiguity in the terms related to delusion, one of Pickering's critics agrees with this diagnosis. Richard Mullen (2003) expressed a similar conclusion to mine:

> None [of the terms] may resolve the questions that Pickering raises, but recognizing that our terms are not always precise or used consistently is part of formulating the problem more clearly. (Mullen, 2003, p.265)

This ambiguity plays a role in the possible actions of the debate as well.

Because the strong objection is more important to the debate about the nature of delusions, and there are, arguably, ways to bypass the weak objection, I will centre the discussion on how the strong objection plays a role in the form of disagreement. The question that arises here is: what is the nature of such disagreement? Is it a merely verbal dispute, such that agreement can be reached once everyone agrees about the terms used, or does it have a different root?

With respect to the strong objection, the lack of theoretical settlement, combined with the lack of consensus, or ambiguity, of the features that constitute delusion given that this is a live debate, leads to overall disagreement. When thinking about the type of disagreement we are faced with in the delusions debate, several possibilities present themselves. The answer depends on the very same nature of the disagreement in the first place.

I suggest the following possibilities for the nature of the disagreement:

1. It indicates a merely verbal dispute. It may be that the differences between two accounts concern merely arbitrary definitions of the features of delusion. Once both sides see that they are merely using the same word for different things - thus generating the appearance of disagreement about the nature of delusion- they can both reach an agreement.

2. There is a substantial, holistic disagreement- Disagreement amongst accounts is principled. In this case, both argue to defend the features of their competing theories. (In the example, above, Campbell has reasons for his Davidsonian approach to the role of rationality and Bortolotti has reasons to disagree with them.) However, those theories depend on other concepts, which are not necessarily about delusions and also are incommensurable. For instance, their theories are dependent on concepts such as functional states, reductionism, or rationality. Therefore, this generates disagreement with respect to this further concept, and so it generates some sort of circularity. No agreement can be reached, and the best that can be claimed is internal consistency. Note that I am not suggesting that internal consistency is easy. Rather, my assertion is that it is not enough to achieve internal consistency to reach a settlement amongst accounts, because two rival accounts might succeed in this. E.g.: in the case of

alcoholism as an illness or as a moral failing, both accounts succeed in internal consistency, so that this alone is not enough to say which account is correct. Equally, doxastic and anti-doxastic accounts might be internally consistent but differ on the right weights to apply to delusional characteristics, therefore being affected by the weak objection. Although, many of my criticisms to particular accounts of delusion point to internal inconsistencies - or at least tensions - within the different rival accounts, the point is that, even if all accounts would solve such internal problems that still will not solve the issue of agreement amongst them.

3. There is a substantial disagreement over a range of theoretical claims but in the end everyone agrees that there is, in principle, a basis to declare which one is the most compelling argument. Eventually everyone will agree on the defining notions of delusion.

So, how should one assess the reach of the strong objection?

Option 1. is not the case because the competing views of delusion disagree in aspects which have deep incommensurable roots other than what the features are called. For example, even if we would all call the patients' tendency to isolate relevant epistemic features as "compartmentalization", the question whether that feature counts towards doxasticism or not, or whether it is a relevant feature at all, or whether it should count for the classification of delusion, would remain untouched and disagreement would persist.

This leaves a final question: could we ever, in the future, arrive at a true definite account of the nature of delusions? The status of the debate is not just a verbal dispute, as in scenario 1., because both sides have principled differences for the reasons I argued above. But, although I acknowledge the difference between option 2. and 3., in practice there is no difference between them. This is so because, given the impact of the likeness argument and its objections, there is not a pragmatic difference: agreement amongst accounts of delusion will not be reached until the major problems of Philosophy have been solved: what is a belief?, what counts as rational?, what are the limits of understandability? In other words, it is not possible to know if the dispute will be solved promptly, or not, given the current status of the debate. This, I have argued, explains the open-ended nature of the disagreements set out in the previous chapters. My

conclusion is, then, that in practical terms, there will not be agreement about the nature of delusions because of the form of argumentation in the debate and its current status.

IX. Importance of the analysis of delusions

In my point of view, there are two main reasons of why an analysis of the debate about delusions is timely and important. The first reason has to do with philosophical thinking whilst the second one has to do with the way clinicians and people in general perceive the phenomenon.

a) Philosophical insight

As it can be seen from the previous chapters, philosophically speaking, it is valuable to have some insight about the assumptions that accounts of delusion take for granted. This questioning is beneficial because it highlights benefits and flaws from different accounts in order to get a better understanding of delusions. The suggestion here is not that philosophy of delusions should be completely abandoned. Mullen (2003) conveys two important considerations that should be taken in account within philosophical debates of mental health in general: that people are beginning to assume themselves with mental health disorders with the purpose of getting help, and also, that clinical psychiatry is being pressured to define as disorders those conditions that have negative social implications. The urgent matter here is that, if philosophers and clinicians do not take a position in this debate, someone else will take the role:

> To a certain extent this brings with itself the risk that social and political fashion may be important determinants of what is conveniently to be considered a disorder, opening the door for all kinds of intolerant and frankly totalitarian approaches. (Mullen, 2003, p.256)

Mullen agrees with the fact that a definition of delusions is inadequate as it is now, but still this is better than to have a false, dogmatic one, “professional recognition that definitions of mental disorder are inadequate, or rather incomplete, may be valid.

Worryingly, there are others who are willing to define it for us" (Mullen, 2003, pp.265-266). He concludes that the highest priority must be to act in the interest of the patient.

b) Clinical significance of delusions

In this work I argue that there is no agreement about what is the correct view about the nature of delusion. So, what are the possible implications for clinical practice given this state of affairs? In my view, what scepticism about delusions might imply is twofold:

1) It might grant freedom for clinicians to seek what might work in a case to case basis because they will no longer be constrained to any assumption in particular; they will no longer think that all delusions are beliefs, or other propositional attitudes, etc.

2) If clinicians do not focus so much on what delusions are, this might encourage them to ask themselves about how the delusion is affecting the person; more explicitly, how can the patients be helped on the basis of harm, not on the basis of having a delusion.

I will explain these two suggestions further.

About 1), if one does not come to a decision about which one is the correct account about delusions, as is the case throughout this work, then the clinical significance is obviously not related to a particular view. As the debate currently stands, I am, in principle, sceptical that we can come to know what delusions are; there are good arguments within all the accounts that are relatively balanced in the sense that all of them have strong advantages, but equally intractable disadvantages too. Furthermore, differences amongst accounts are necessarily part of the debate, as I argue in this section. So, one of my conclusions is that no assumption is "the correct one", and so, clinicians have the option of being undogmatic about treatment.

For instance, if we just knew that delusions were beliefs, then that would give us a particular reason for thinking that talking therapies would be an appropriate treatment. If we, instead, decided that the importance about delusions was not the belief's content, but about something such as the phenomenological delusional atmosphere, then we might consider a different treatment approach. So, it seems that one of the implications from my thesis is that we should stop endorsing any one of these assumptions; there is

a kind of plurality of theoretical perspectives about what delusions are, and this opens up the possibility of granting clinicians the freedom to try to see what works in particular cases.

About 2), perhaps what matters is not what delusions are, but the kind of distress that they may cause. Given my position of being sceptical about the nature of delusion, one prospect could be to focus, not on what a delusion is, but instead, on what is more relevant to patients, that is, the distress they are going through.

GENERAL CONCLUSION

In this work I have presented an analysis of some of the most representative contemporary philosophical accounts of delusion. I have made explicit their key underlying assumptions and I have criticised the resulting accounts. In the last chapter I offer a diagnosis about the patent disagreement amongst these accounts.

In the first chapter, I developed the concepts needed to arrive at the two main definitions of delusions suggested by Jaspers: the first one regarding the characteristics of delusion, and the second one that portrays delusions as the ununderstandable. The first definition led to modern definitions in the DSM and elsewhere, whilst the second one was regarded by Jaspers as the better, more fundamental way to think about delusions. The idea that delusions are ununderstandable has been considered unsatisfactory to modern philosophers. One of the reasons is that there seems to be a tension in Jaspers in that "understanding" is at the heart of his psychiatry in all of its concepts, except for delusions. This tension is bigger if one thinks that being understandable is a mark of all the other mental states. Does that mean that delusions should not be considered as mental states? This does not seem plausible. The implicit hope of recent philosophical work on the nature of delusions seems to be that, by saying what sort of state a delusion is, light can be shed on their status as understandable or not.

In the second chapter I analysed antidoxastic accounts of delusion. Antidoxasticism responds to the worry that delusions cannot be beliefs by suggesting that they are some other non-doxastic propositional attitude. Their main advantage is having explanatory power, but at the cost of being too ad hoc. Antidoxastic alternatives have been unable to refute the main claims of doxasticism, or to give a solution to all the problems encountered in descriptions of delusion. Their strategy faces the following dilemma:

1) Either delusions are compared to a familiar propositional attitude- If so, this promises to shed light on delusions, because we are familiar with the propositional attitude in question. However, there are significant ways in which delusions seem unlike familiar propositional attitudes, which poses a serious challenge to this approach.

2) Or they compare delusions to non-standard propositional attitudes- But now the objection is that the ad hoc nature of the propositional attitude means no light is shed on delusion because we do not understand what it would be to have such a

propositional attitude. For example, a bimagining; a delusional stance; a cognitive hallucination etc. This means that we end up with something sui generis that is not readily understood. Thus antidoxasticism faces fundamental difficulties in responding to the challenge implicitly set by Jaspers.

In the third chapter I looked at doxastic accounts of delusion. Philosophers have been attracted to this stance because delusions are held with strong conviction, and they seem to have many characteristics of normal beliefs. However, the problem with doxastic accounts is that there is not a definitive answer to two fundamental aspects: What are the fundamental constitutive elements of a belief? The different accounts are forced to produce a flexible concept of belief. It is clear that a conventional view of what constitutes a belief cannot be applied straightforwardly to a controversial subject such as delusions. One can also observe that belief accounts are also somewhat incompatible among doxastic views because their guiding assumptions are different. A second aspect is that there is no agreement because the existing accounts cannot accommodate the bizarre characteristics of delusion to an account that will work for all the other non-delusional beliefs, even when the concept is more flexible. In the end, the fundamental problem of doxasticism is that one starts with the idea that delusions do not seem like ordinary beliefs: they are obviously quite odd. Doxasticists then argue that, despite those appearances, they are enough like beliefs to count as beliefs. So that leaves the question, why do they seem unlike beliefs? The more doxasticism does to show that they are akin to beliefs, the harder it will be to account for their manifest strangeness.

The fourth chapter focuses on phenomenological accounts, which differ greatly from the ones in the previous chapters. Although the same empirical data has to be explained, the importance given to the facts differs. In phenomenological accounts the fundamental question is: What is the nature of the patient's experience? 'Experience' is not, however, meant as a definite set of measurable facts. Rather, it is the opposite view. The patient's subjective experience is what has to be taken into account. This emphasis on subjectivity makes phenomenological accounts attractive to patients as a description of their illness; its richness is something worth analysing as a possible answer to the nature of delusions in the first place. In the end, however, phenomenological stances do not give a satisfactory account of the nature of delusions. The main reason for this comes from the vagueness of the explanatory devices such as "alternative realities", or "radical empathy". Such concepts do not explain the main problematic features of

delusions, like their pathological nature. In other words, phenomenologists play down the problems encountered in delusions to give an account that encompasses the subjective nature rather than the specific puzzling questions of delusions. Therefore, although such accounts may fit patients' narratives, the initially pathological and puzzling nature remains.

The fifth chapter considered the question about the meaning or content of delusions. Three main theories were considered: delusions as empty speech acts, delusions as framework propositions, and delusions as quasi-solipsism. The former account, delusions as empty speech acts, suggests that delusions have no meaning and their content is just random. The two latter accounts try to "save" the meaning of delusions by proposing that they are akin to something familiar, although the familiar part is also strange. These three accounts try to preserve the sense in which delusions are bizarre. At the same time, these authors are trying to address the apparent difficulties in understanding what patients think or mean. However, this leads to a familiar difficulty: patients seem to understand what they say, but the link to rationality seems to be lost. So, in the end, these three accounts cannot explain away the problems about meaning or content in delusions.

Summarizing, the thesis has argued that none of the various theories of the nature of delusions succeeds. All accounts face different objections. However, the general problem is that the accounts all attempt to shed light by comparing delusions to more familiar kinds of mental states. Still, this faces the difficulty of making delusions too easily understandable. Retreating from this by defining more bespoke, less familiar mental states threatens to make the accounts ad hoc and unilluminating. My final aim was an attempt to show why this is so: there is something problematic about the form of analysis offered.

In the final chapter the debate about delusions was compared to Pickering's (2003) critique of the "likeness argument" concerning mental illness. I argued that accounts of the nature of delusions take the form of a likeness argument by arguing that delusions belong to a particular category of mental states because they share enough of the characteristics of paradigm cases of that category. If this is so, then they face two principled objections: the weak and strong objections to the grounding assumptions of the likeness argument, provided that there is an ongoing debate. The weak objection argues that, in the case of delusions, the weight assigned to the description of the

features depends on the theory endorsed. The strong objection argues that the particular framework used by each author will influence the features or characteristics assigned to the phenomenon, therefore begging the question from the beginning. Because the all the accounts analysed through this work have the form of the likeness argument, they cannot avoid assuming a priori some of the features they want to explain, and this presumes what it is, instead, supposed to prove.

My general conclusion is that delusions cannot be non-question-beggingly described by any philosophical theory because they fall prey to the likeness argument's objections. In the case of delusions one could adopt the position that delusions are a sui generis phenomenon, something unique that is not like anything else. This approach would be similar to Jaspers' definition of delusions as "the ununderstandable". However, such a position is surely unsatisfying due to the combination of delusions having many recognizable features and yet remaining prima facie puzzling.

Although the latter alternative is unsatisfying, it is hard to see how the philosophical project of trying to shed light on the nature of delusions can avoid the strategy of comparing them to more familiar mental states or propositional attitudes.

Given my position of being sceptical about the nature of delusion, one prospect could be to focus, in the future, not on what a delusion is, but instead, on what is more relevant to patients, that is, the distress they are going through.

REFERENCES

Ajzen, I. (2005). *Attitude, personality and behaviour*. Maidenhead: McGraw-Hill.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

Ames, D. (1984). Self shooting of a phantom head. *British Journal of Psychiatry*, 145: 193–194.

Aronson, E. (1999). Dissonance, hypocrisy and the self-concept*. In E. Harmon-Jones and J. Mills (Eds.) Cognitive Dissonance: Progress in a pivotal theory in social Psychology. Washington DC, APA,* pp. 103-126.

Austin, J. L. (1962). How to do things with words (JO Urmson, Ed.). New York: Oxford University.

Basu, A., Singh, P., Gupta, R., Kundu, S. (2013). Electroconvulsive therapy for long-term mutism in a case of noncatatonic paranoid schizophrenia. Innovations In Clinical Neuroscience, 10(7-8), 10-12.

Bauer R. M. (1984). Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the Guilty Knowledge Test. *Neuropsychologia 22* (4): 457–69.

Bauer R. M. (1986). The cognitive psychophysiology of prosopagnosia*. In H. G. Ellis, M.A. Jeeves, F. Newcombe and A. Young (Eds.), Aspects of face processing, Dordrecht, Martinus, Nijhoff*, pp. 253-267.

Bayne, T. (2010). Delusions as doxastic states: Contexts, compartments, and commitments. Philosophy, Psychiatry, & Psychology, 17(4), 329-336.

Bayne, T., Pacherie, E. (2004). Bottom-Up or Top-Down: Campbell's Rationalist Account of Monothematic Delusions. *Philosophy, Psychiatry, and Psychology 11* (1):1-11.

Bayne, T., Pacherie, E. (2005) In Defence of the Doxastic conception of Delusions. *Mind and Language, 20*, pp.163-188.

Bentall, P.P. (1994) Cognitive biases and abnormal beliefs: Towards a model of persecutory delusions. *In A. S. David and J.C. Cuttieng (eds), The Neuropsychology of Schizophrenia. Hove, E. Sussex: Psychology Press*, 337-360.

Bentall, R.P., Kaney, S. (1996). Abnormalities of self-representation and persecutory delusions: A test of a cognitive model of paranoia. *Psychological Medicine, 26*, 1231-1237.

Berrios G.E., Luque R., (1995). Cotard's Delusion or Syndrome?: A Conceptual History, *Comprehensive Psychiatry, Vol. 36, No. 3* (May/June), pp 218-223.

Berrios, G. E. (1991). Delusions as "wrong beliefs": a conceptual history. The British Journal of Psychiatry. 159 (Suppl.14) 6-13.

Berrios, G. E. (1996). The history of mental symptoms: descriptive psychopathology since the nineteenth century. Cambridge University Press.

Bleichrodt, H., Pinto Prades, J. (2009). New evidence of preference reversals in health utility measurement, *Health Economics, John Wiley & Sons, Ltd., vol. 18*(6), pp 713-726.

Bleuler, E. P. (1906). Affektivität, Suggestibilität, Paranoi. Halle.

Bleuler, E. P. (1950). Dementia Praecox or the group of Schizophrenias.

Blount, G. (1986). Dangerousness of patients with Capgras syndrome. *Nebraska Medical Journal* 71: 207.

Bloy, S., Morris, E., Oliver, J. (2011). Using acceptance and commitment therapy with people with psychosis: A case study. Clinical Case Studies, 10(5), 347-359.

Boorse, C. (1977). Health as a theoretical concept. Philosophy of Science 44:542–573.

Boorse, C. (1982). What a theory of mental health should be. Ed. R. Edwards. Psychiatry and ethics: Insanity, rational autonomy, and mental health care (pp. 29–48). Buffalo, NY: Prometheus Books.

Bortolotti, L. (2005). Delusions and the Background of Rationality. *Mind and Language, 20,* pp.189-208

Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford University Press, Oxford

Bortolotti, L. (2012). In Defence of Modest Doxasticism about Delusions. *Neuroethics 5* (1):39-53.

Bortolotti, L. (2012a). Précis of Delusions and Other Irrational Beliefs. *Neuroethics* 5 (1):1-4.

Bortolotti, L. (2012, June 14). Self-narratives and autonomy. Retrieved November 06, 2016, from http://kolber.typepad.com/ethics_law_blog/2012/06/self-narratives-and-autonomy-by-lisa-bortolotti.html

Breen, N., Caine, D., Coltheart, M., Hendy, J., Roberts, C. (2000) Towards an Understanding of Delusions of Misidentification: Four Case Studies, *In Pathologies of Belief, Coltheart and Davies (Eds.), Blackwell,* pp.75-110.

Brett-Jones, J., Garety, P., Hemsley, D. (1987). Measuring delusional experiences: A method and its application. *British Journal of Clinical Psychology*, 26, 257–265.

Brody, H. (1985). Philosophy of medicine and other humanities: toward a wholistic view. Theoretical Medicine 6:243–255.

Brown, W. (1985). A critique of three conceptions of mental illness. The Journal of Mind and Behavior 6:553–576.

Cahill, C., Frith, C. (1996). False perceptions or false beliefs? Hallucinations and delusions in schizophrenia*. In Method in Madness: Case Studies in Cognitive Neuropsychiatry (eds P.W. Halligan & J.C. Marshall),* pp. 267-291. Hove: Psychology Press.

Campbell, J. (1999). Schizophrenia, the space of reasons and thinking as a motor process. *The Monist 82* (4):609-625.

Campbell, J. (1999a). Immunity to error through misidentification and the meaning of a referring term. *Philosophical Topics 26* (1/2):89-104.

Campbell, J. (2001). Rationality, meaning, and the analysis of delusion. *Philosophy, Psychiatry, and Psychology 8* (2-3):89-100.

Campbell, J. (2008). Causation in psychiatry. *In: K.S. Kendler and J. Parnas (eds). Philosophical issues in psychiatry. Baltimore: Johns Hopkins University Press.*

Candido, C. L., Romney, D. M. (1990). Attributional style in paranoid vs. depressed patients. *British Journal of Medical Psychology, 63*: 355–363.

Capgras, J.; Reboul-Lachaux, J. (1923). Illusion des "sosies" dans un délire systématisé chronique. *Bulletin de la Société Clinique de Médicine Mentale 2*: 6–16.

Carruthers, P. (2005). *Consciousness: Essays From a Higher-Order Perspective*, Oxford: Clarendon Press.

Chadwick, P.D.J., Lowe, C.F. (1990). Measurement and modification of delusional beliefs. *Journal of Consulting and Clinical Psychology, 58*, 225-232.

Cherniak, C. (1986). *Minimal Rationality*. Cambridge, MA: MIT Press.

Chinn, C. A., Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction, 19*, 323–393.

Claridge, G. (1992). Can a disease model of schizophrenia survive? Ed. R. Bentall. Reconstructing schizophrenia (pp. 157–183). London: Routledge.

Coltheart, M., Langdon, R., McKay, R. (2007). Schizophrenia and monothematic delusions*. Schizophrenia bulletin, 33*(3), 642-7.

Coltheart, M., Menzies, P., et al. (2010). Abductive inference and delusional belief. *Cognitive Neuropsychiatry, 15*(1), 261.

Conrad, K. (1958). Die beginnende Schizophrenie: Versuch einer gestaltanalyse des Wahns [The onset of schizophrenia: An experimental analysis of creative madness], Sttutgart, Thieme.

Cortlett, P.R., Krystal. J.H., Taylor, Fletcher, P.C. (2009) Why do Delusions Persist*? Frontiers of Human Neuroscience, 3* (12), p. 1-9.

Cortlett, P.R., Taylor, J.R., Wang, X. J., Fletcher, P.C., Krystal. J.H. (2010) Toward a Neurobiology of Delusions*. Progress in Neurobiology 92*, pp.345-369.

Cotard J. (1880). Du delire hypocondriaque dans une forme grave de la melancolie anxieuse. *Ann Med Psychol* (Paris); 4:168-174.

Crider, A. (1997) Perseveration in Schizophrenia. Schizophrenia Bulletin, Vol. 23, No. 1, 62-74.

Culver, C., B. Gert. (1982). Philosophy in medicine. New York: Oxford University Press.

Currie, G. (2000). Imagination, delusion, and hallucination. *Mind & Language 15*, pp.168–83.

Currie, G., Jureidini, J. (2001). Delusion, rationality, empathy. *Philosophy, Psychiatry & Psychology, 8/2,3*, 159–62.

Currie, G. and Ravenscroft, I. (2002). *Recreative Minds*. Oxford: Oxford University Press.

Davidson, D. (1973). On the very idea of a conceptual scheme. *Proceedings of the American Philosophical Association,* 47, 5–20.

Davidson, D. (1982). Paradoxes of irrationality*. In Philosophical essays on Freud, ed. R. Wolheim and J. Hopkins*, 289–305. Cambridge: Cambridge University Press.

Davidson, D. (1984). *Inquiries into truth and interpretation*. Oxford: Oxford University Press.

Davies M., Coltheart M., (2000). Introduction*. Pathologies of belief. Blackwell publishers*. p. 1-46.

Davies, A.M., Davies, M., Ogden, J.A., Smithson, M., and White, R.C. (2009). Cognitive and motivational factors in anosognosia. *In T. Bayne, and J. Fernández (eds.) Delusions and Self-Deception: Affective and Motivational Influences on Belief Formation*, pp. 187–225. Hove, East Sussex: Psychology Press.

Davies, M., Coltheart M., Langdon, R., Breen, N., (2001). Monothematic delusions: Towards a Two-Factor account, *Philosophy, Psychiatry and Psychology,* 8, pp. 133-158.

de Clerambault, G. D. (1909). Oeuvre. Tome I, 145-210.

de Clerambault, G. D. (1942). Les psychoses passionelles (1921*). In Oeuvre Psychiatrique. Paris: Presses Universitaries de France.*

Debruyne, H., Portzky, M., Van den Eynde, F., Audenaert, K. (2009). Cotard's syndrome: a review. Current psychiatry reports, 11(3), 197-202.

Dennett, D.C. (1971) Intentional systems. *Journal of Philosophy, 8*, 87–106.

Di Tella, R., Galiani, S., Schargrodsky, E. (2004). Property rights and beliefs: evidence from the allocation of land titles to squatters. Harvard Business School mimeo.

Doerr-Zegers, O., Pelegrina-Cetrán (2013). Karl Jaspers' General psychopathology in the framework of clinical practice. *In G. Stanghellini, & Th. Fuchs. One century of Karl Jaspers' General psychopathology* (p. 58-75). Oxford University Press.

Donagan, A. (1978). How much neurosis should we bear? Ed. H. Engelhardt Jr. and S. Spicker. Mental health: Philosophical perspectives (pp. 41–53).

Egan, A., (2008), Imagination, Delusion, and Self-Deception, *in Delusion and Self-Deception: Affective Influences on Belief-formation, Bayne and Fernandez (eds.), Psychology Press,* pp.263–280.

Eilan, N. (2000). On understanding schizophrenia*. In D. Zahavi (Ed.), Exploring the self* (pp. 97–113). Amsterdam: John Benjamins.

Ellis H.D., Young A.W., Quayle A.H., de Pauw, K. W. (1997) Reduced autonomic responses to faces in Capgras delusion. *Proc R Soc Lond Biol Sci*.;p.264:1085–1092.

Ellis, H. D., Young, A. W. (1990). Accounting For Delusional Misidentifications*. British Journal of Psychiatry,* 157, 239-248.

Ellis, H. D., Lewis, M.B. (2001). Capgras delusion: A window on face recognition. *Trends in Cognitive Sciences* 5: 149–156.

Feinberg, T. E., Eaton, L. A., Roane, D. M., Giacino, J. T. (1999). Multiple Fregoli delusions after traumatic brain injury. *Cortex: A Journal Devoted To The Study Of The Nervous System And Behavior, 35*(3), 373-387.

Fine C., Gardner M., Craigie J., Gold I. (2007) Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cogn. Neuropsychiatry*.;1:46–77.

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., Combs, B. (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. Policy sciences, 9(2), 127-152.

Flew, A. (1973). Crime or disease? New York: Barnes and Noble.

Forstl, H., Almeida, O. P., Owen, A. M., Burns, A. and Howard, R. (1991): Psychiatric, neurological, and medical aspects of misidentification syndrome: A review of 260 cases. *Psychological Medicine*, 21, 905–10.

Frankish, K. (2009). Delusions: A two-level framework. *In Matthew Broome & Lisa Bortolotti (eds.), Psychiatry as Cognitive Neuroscience: Philosophical Perspectives. Oxford University Press.* p.269—284.

Frankish, K. (2012). Delusions, Levels of Belief, and Non-doxastic Acceptances. *Neuroethics*, 23-27.

Freeman, D., Garety, P. A. (2003). Connecting neurosis and psychosis: The direct influence of emotion on delusions and hallucinations. *Behaviour Research and Therapy*, 41(8), 923–947.

Freeman, D., Garety, P. A., Kuipers, E., Fowler, D., Bebbington, P. E. and Dunn, G. (2007). Acting on persecutory delusions: the importance of safety seeking. *Behaviour Research and Therapy*, 45, 89–99.

Frith, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Psychology Press.

Fulford, K. (1989). Moral Theory and Medical Practice. Cambridge: Cambridge University Press.

Fulford, K.W.M, Thornton T., Graham, G. (2006*). Oxford Textbook of Philosophy and Psychiatry.* Oup Oxford.

Fulford, K.W.M., (1998). Completing Kraepelin's psychopathology: Insight, delusion and the phenomenology of illness, *in X. F. Amador and A. David (eds.) Insight and Psychosis, Oxford: Oxford University Press:* 47–65.

Gallagher, S. (2004a). Neurocognitive models of schizophrenia: A neurophenomenological critique. *Psychopathology*, 37, 8–19.

Gallagher, S. (2004b). Agency, ownership and alien control in schizophrenia. *In The Structure and Development of Self-consciousness: Interdisciplinary Perspectives (eds. P. Bovet, J. Parnas, and D. Zahavi),* pp 89–104. John Benjamins, Amsterdam.

Gallagher, S. (2009). 'Delusional realities'*. in M Broome & L Bortolotti (eds), Psychiatry as Cognitive Neuroscience: Philosophical Perspectives. International Perspectives in Philosophy & Psychiatry, OUP,* pp. 245-66.

Gerrans, P. (2000). Refining the explanation of Cotard's Delusion, *in Davies M., Coltheart M., (Eds). Pathologies of belief. Blackwell publishers.* pp. 111-122.

Gerrans, P. (2012). Dream experience and a revisionist account of delusions of misidentification. *Consciousness and Cognition* (21), pp. 217-227.

Gibbs, A. A., David, A. S. (2003). Delusion formation and insight in the context of affective disturbance. Epidemiologia e psichiatria sociale, 12(03), 167-174.

Gipps, R. G. (2012). The indefinability and unintelligibility of delusion. Philosophy, Psychiatry, & Psychology, 19(2), 91-95.

Gipps, R. G. T., Fulford, K. W. M. (2004). Understanding the Clinical Concept of Delusion. From an Estranged to an Engaged Epistemology, International Review of Psychiatry, 3: 225-235.

Gold, I. and Hohwy, J. (2000). Rationality and schizophrenic delusion. *In Pathologies of Belief (eds. M. Coltheart, and M. Davies)*, pp 145–165. Blackwell, Oxford.

Graham, G., (2010) Are the Deluded Believers? Are Philosophers among the Deluded*? Philosophy, Psychiatry, and Psychology,* 17.4: pp.337-339.

Griesinger, W. (1845). (English trans, of 2nd edn., 1867). Die Pathologie und Thérapie der psychischen Krankheiten. Stuttgart.

Hagen, F. W. (1870). Studien auf dem Gebiete der aerztlichen Seelenkunde. Erlangen: Besold.

Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley.

Hemsley, D. R., Garety, P. A. (1986). The formation of maintenance of delusions: a Bayesian analysis. The British Journal of Psychiatry, 149(1), 51-56.

Hertwig, R.; Gigerenzer G., (1999). The 'Conjunction Fallacy' Revisited: How Intelligent Inferences Look Like Reasoning Errors. *Journal of Behavioral Decision Making* 12: 275–305.

Hoerl, C. (2013). Jaspers on explaining and understanding in psychiatry. *In Thomas Fuchs & Giovanni Stanghellini (eds.), One Hundred Years of Karl Jaspers' General Psychopathology*. p.107-120. Oxford University Press.

Hohwy, J. (2004). Top-down and bottom-up in delusion formation. *Philosophy Psychiatry and Psychology 11* (1):65-70.

Hohwy, J. (2007). The Sense of Self in the Phenomenology of Agency and Perception, *Psyche 13* (1), pp. 1-20.

Hohwy, J., and V. Rajan. (2011). Delusions as forensically disturbing perceptual inferences. *Neuroethics 5* (1):5-11.

Howe, G. (1995). Working with Schizophrenia: A Needs Based Approach. Jessica Kingsley Publishers.

Huq, S. F., Garety, P. A., Hemsley, D. R. (1988). Probabilistic judgements in deluded and non-deluded subjects. *Quarterly Journal of Experimental Psychology*, 40A, 801-812.

Jaspers, K. ([1913] 1997). *General Psychopathology,* trans. J. Hoenig & M.W. Hamilton. Baltimore, MD: Johns Hopkins University Press.

Jaspers, K. (1913a). Causal and "Meaningful" Connections between Life History and Psychosis. *In: Kirsch, S.R.; Shepherd, M. (eds). Themes and Variations in European Psychiatry. Bristol: Wright & Sons*. pp. 80-93.

Jaspers, K. (1968)[1912] The phenomenological approach in psychopathology. *British Journal of Psychiatry;* 114, pp.1313-1323.

Kapur, S. (2003). Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. *American Journal of Psychiatry*, 160, pp.13–23.

Kapusta, A. (2010). Madness and Method: The limit of understanding in Philosophy and Psychiatry, Lublin: UMCS

Kemp, R., Chua, S., McKenna, P., David, A. (1997). Reasoning and delusions. The British Journal of Psychiatry, 170(5), 398-405.

Kepska, A., Hawro, T., Krupinska-Kun, M., Zalewska, A. (2011). Somatic-type delusional disorder: a case report and comments. Acta Dermato-Venereologica, 91(2), 193-194.

Kinderman, P., Bentall, R.P. (1997). Causal attributions in paranoia: Internal, personal and situational attributions for negative events. *Journal of Abnormal Psychology,* 106, 341-345.

Komiyama, M. (1989). Fictionality of schizophrenic delusions. Psychiatry and Clinical Neurosciences, 43(1), 13-18.

Kopelman, M.D. (1987). Two types of confabulation. *J Neurol Neurosurg Psychiatry* a; 50: 1482–7.

Kovács, A., Vörös, V., & Fekete, S. (2005). Suicide attempt and melancholic depression in a male with erotomania: case report. Archives of Suicide Research, 9(4), 369-372.

Kraus, A., (2014), Karl Jaspers on Primary Delusional Experiences of Schizophrenics: His Concept of Delusion Compared to That of the DSM. *In Thomas Fuchs, Thiemo Breyer , Christoph Mundt, (eds.) Karl Jaspers' Philosophy and Psychopathology, Springer*, p. 109-124.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lagnado, D. A., Sloman, S. A. (2006). Time as a guide to cause. *Journal of experimental psychology. Learning, memory, and cognition,* 32(3), 451–60.

Lai, D. C., Chang, W. C., Tam, W. Y., Hui, C. M., Chen, E. H. (2013). Cognitive and affective perspectives on formation and maintenance of grandiose delusions of a patient with schizophrenia. East Asian Archives of Psychiatry, 23(4), 160-163.

Loughlin, M. (2003). Contingency, arbitrariness, and failure. Philosophy, Psychiatry, & Psychology, 10:261–264.

Lucchelli, F., Spinnler, H., (2007). The case of lost Wilma: a clinical report of Capgras delusion, *Neurological Science,* No. 28, pp. 188–195.

Lukes, S. (1982). Relativism in its place. *In Martin Hollis & Steven Lukes (eds.), Rationality and Relativism. MIT Press*. 261--305.

Maher, B. A. (1974). Delusional thinking and perceptual disorder. *Journal of Individual Psychology*, 30, 98–113.

Marco-Merenciano, F. (1942) Esquizofrenias paranoides. Madrid, Spain: Miguel Server.

McDowell, J. (1978). Are moral requirements hypothetical imperatives? *Proceedings of the Aristotelian Society* 52(suppl.):13–29.

McDowell, J. (1994). *Mind and world.* Cambridge, MA: Harvard University Press.

McGinn, M. (1989). *Sense and certainty*. Oxford: Blackwell.

Mele, A.R. (1993). Motivated belief. *Behaviour & Philosophy*, 21/2, 19–27.

Miller, D.; Ross M., (1975). Self-serving Biases in the Attribution of Causality: Fact or Fiction?. *Psychological Bulletin 82* (2): 213–225.

Mishara, A., Fusar-Poli, P. (2013). The phenomenology and neurobiology of delusion formation during psychosis onset: Jaspers, Truman symptoms, and aberrant salience. Schizophrenia Bulletin, 39(2): 278-286.

Moore, G. E. (1939). Proof of an external world. *Proceedings of the British Academy* 25:273–300.

Mullen, R. (2003). Definition is limited and values inescapable. Philosophy, Psychiatry, & Psychology, 10:265–266.

Murphy, D. (2012). The folk epistemology of delusions. *Neuroethics 5* (1):19-22.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review 83* (October):435-50.

Nisbett, R. E., Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological review, 84(3), 231.

Nozick, R. (2001). *Invariances: The Structure of the Objective World*. Belknap Press of Harvard University Press.

O'Connor, A. R., Moulin, C. J. A. (2010). Recognition without identification, erroneous familiarity, and déjà vu. *Current Psychiatry Reports*, 12(3), 165–173.

Parnas, J., Sass, L. (2001). Self, solipsism, and schizophrenic delusions. *Philosophy, Psychiatry, and Psychology,* 8(2/3),pp.101–120.

Pickering, N. (2003). The Likeness Argument and the Reality of Mental Illness. *Philosophy, Psychiatry, and Psychology 10* (3):243-254.

Pickering, N. (2003a). The likeness argument: Reminders, roles, and reasons for use. Philosophy, Psychiatry, & Psychology, 10(3), 273-275.

Pickering, N. (2006). The Metaphor of Mental Illness. Oxford University Press. UK

Price, H.H., (1934). Some considerations about belief. Proceedings of the Aristotelian Society, 35, 229-252.

Quattrone G.A. and Tversky A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology* 46 (2): 237-248.

Quine, W.V.O. (1960). *Word and Object.* Cambridge, MA: MIT Press.

Radden, J. (2006). Review - The Metaphor of Mental Illness. Retrieved June/July, 2016, from http://metapsychology.mentalhelp.net/poc/view_doc.php?type=book&id=3162&cn=458

Ramachandran, V. S., and Blakeslee, S. (1998). *Phantoms in the brain: Probing the mysteries of the human mind* (1st ed.). New York: William Morrow.

Ramsey, F.P. (1931). General Propositions and Causality, *in R.B. Braithwaite ed. F.P. Ramsey: The Foundations of Mathematics* (Routledge and Kegan Paul, London, p.238.

Ratcliffe, M. (2008). The phenomenological role of affect in the Capgras delusion. Continental Philosophy Review, 41(2), 195-216.

Ratcliffe, M. (2010). Delusional Atmosphere and Delusional Belief. *In Handbook of Phenomenology and Cognitive Science. Gallagher, S. & Schmicking, D. Dordrecht: Springer.* pp. 575-590.

Ratcliffe, M. (2013). Delusional Atmosphere and the Sense of Unreality. *In One Century of Karl Jaspers' General Psychopathology. Stanghellini, G. & Fuchs, T. Oxford: Oxford University Press*. 229-244.

Ratcliffe, M. (2014). The phenomenology of depression and the nature of empathy. *Medicine, Health Care and Philosophy, Volume* 17, Issue 2, pp. 269-280.

Read, R J. (2003). Literature as Philosophy of Psychopathology: William Faulkner as Wittgenstein. *Philosophy, Psychiatry, and Psychology 10* (2):115-124.

Read, R. J. (2003a). On Delusions of Sense: A Response to Coetzee and Sass. *Philosophy, Psychiatry, and Psychology 10* (2):135-141.

Régis, E., Hesnard, A. (1924). La Psychoanalyses des Nevroses et des Psychoses, ses Applications Medicales et Extra-medicales. The Journal of Nervous and Mental Disease, 59(1), 105.

Reimer, M. (2010). Only a Philosopher or a Madman: Impractical Delusions in Philosophy and Psychiatry. *Philosophy, Psychiatry, and Psychology 17* (4), pp.315-328.

Reimer, M. (2010a). Distinguishing Between the Psychiatrically and Philosophically Deluded: Easier Said Than Done. *Philosophy, Psychiatry, and Psychology 17* (4):341-346.

Revonsuo, A. (2006). Inner presence: Consciousness as a biological phenomenon. MIT Press.

Revonsuo, A., Tarkko, K. (2002). Binding in dreams – The bizarreness of dream images and the unity of consciousness. *Journal of Consciousness Studies, 9*(7), 3–24.

Rosset, E. (2008). It's no accident: Our bias for intentional explanations. *Cognition 108*, 771-780.

Samuels, S., Stich, S. Bishop M. (2002). Ending the rationality wars: How to make disputes about human rationality disappear. *In Elio (ed.), Common Sense, Reasoning and Rationality. Oxford University Pres*s 236-268.236-68.

Sass, L. A. (1994). *The Paradoxes of Delusion: Wittgenstein, Schreber, and the Schizophrenic* Mind. Ithaca, N.Y.: Cornell University Press.

Sass, L. A. (1994a). "My So-Called Delusions": Solipsism, Madness, and the Schreber Case. *Journal of Phenomenological Psychology 25* (1):70-103.

Sass, L. A. (2001). Self and world in schizophrenia: Three classic approaches. Philosophy, Psychiatry, & Psychology, 8(4), 251-270.

Sass, L. A. (2003). Incomprehensibility and Understanding: On the Interpretation of Severe Mental Illness. *Philosophy, Psychiatry, and Psychology 10* (2):125-132.

Sass, L. A. (2014) Delusion and double Book-Keeping In Thomas Fuchs, Thiemo Breyer , Christoph Mundt, (eds.) Karl Jaspers' Philosophy and Psychopathology, Springer, p. 125-147.

Scarone, S., M. Manzone, O. Gambini, I. Kantzas, I. Limosani, A. D'Agostino, and J. A. Hobson. (2008). The dream as a model for psychosis: An experimental approach using bizarreness as a cognitive marker. *Schizophrenia Bulletin 34*, no.3:515–22.

Schneider, K., (1959) General Psychopathology. New York, NY: Grune & Stratton.

Schreber, D. P. (1903). *Memoirs of My Nervous Illness.* New York: New York Review of Books, 2000.

Schwitzgebel, E. (2002): A phenomenal, dispositional account of belief. *Nouˆs*, 36/2, 249–275.

Silva, A., Leong, G. B., Weinstock, R. , Sharma, K. K., and Klein, R. L. (1994) Delusional misidentification syndromes and dangerousness, *Psychopathology, vol*. 27, no. 3–5, pp. 215–219, 1994.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. Annual review of psychology, 28(1), 1-39.

Stalmeier, P.F.M, Wakker, P.P , Bezembinder, T.G.G (1997), Preference reversals: violations of unidimensional procedure invariance, Jo*urnal of Experimental Psychology: Human Perception and Performance, 23*, pp. 1196–1205.

Stanghellini, G. (2004*). Disembodied spirits and deanimated bodies*. Oxford: Oxford University Press.

Stanghellini, G. (2009). Schizophrenic Delusions, Embodiment and the Background. Philosophy, Psychiatry & Psychology, 15(4): 311-314.

Stein, E., (1996). *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science.* Oxford University Press.

Stephens, G.L., Graham, G. (2004). Reconceiving delusion. *International Review of Psychiatry* 1, pp.236–41.

Stephens, G.L., Graham, G. (2007) The delusional stance *in Reconceiving Schizophrenia by Man Cheung Chung, K.M.W.. Fulford, George Graham (Eds.). Oxford: Oxford University Press*, pp.193-215.

Stich, S. P. (1981). Dennett on intentional systems. *Philosophical Topics 12* (1):39-62.

Stich, S. P. ([1984] 2012). Collected Papers, Volume 2: Knowledge, Rationality, and Morality, 1978-2010. Oxford University Press.

Stone, M., Young, A. W. (1997). Delusions and brain injury: The philosophy and psychology of belief. *Mind and Language,* 13, 327–364.

Störring, G. (1939). *Wesen und Bedeutung des Symptoms der Ratlosigkeit bei psychischen Erkrankungen:(ein Beitrag zur Differentialdiagnose der Geistesstörungen)*. Thieme.

Strawson, P. F. (1958). Persons. *Minnesota Studies in the Philosophy of Science* 2:330-53.

Szasz, T. (1960). The myth of mental illness. American Psychologist 15:113–118.

Thornton, T. (2007). *Essential Philosophy of Psychiatry*. Oxford University Press.

Thornton, T. (2008). Why the idea of framework propositions cannot contribute to an understanding of delusions. *Phenomenology and the Cognitive Sciences 7* (2):159-175.

Thornton, T. (2012). Delusional Atmosphere, the Everyday Uncanny, and the Limits of Secondary Sense. *Emotion Review 4* (2):192-196.

Thornton, T. (2015). Against Explanatory Minimalism in Psychiatry. *Front. Psychiatry* 6:171, pp.1-9

Tranel, D., Damasio, H., and Damasio A. R., (1995) Double Association between over and covert recognition. *Journal of Cognitive Neuroscience*, 7, 425-432.

Tumulty, M. (2011). Delusions and Dispositionalism about Belief. *Mind and Language 26* (5):596-628.

Tumulty, M. (2012). Delusions and Not-Quite-Beliefs. Neuroethics 5 (1):29-37.

Tversky, A., Kahneman, D. (1981). The Framing of decisions and the psychology of choice. *Science 211* (4481): 453–458.

Tversky, A., Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90, 293–315.

Tversky, A., Thaler, R., (1990). Anomalies: Preference Reversals. *Journal of Economic Perspectives,* 4(2): 201-211.

Tyreman, S. (2003). Likening strikes twice: psychiatry, osteopathy and the likeness argument. Philosophy, Psychiatry, & Psychology 10:267–271.

Unger, P. (1979). I do not exist. *In Perception and identity, ed. G. F. MacDonald*, 235–51. Ithaca, NY: Cornell University Press.

Vaxevanis A., Vidalis A, (2005), Cotard's syndrome. A three-case report. *Hippokratia,* 9, 1: 41-44.

Venneri, A., Shanks, (2004) M. F., Belief and awareness: reflections on a case of persistent anosognosia, *Neuropsychologia* 42, 230–238.

Wakefield, J. (1992a). The concept of mental disorder. On the boundary between biological facts and social values. American Psychologist 47:373–388.

Wegner, D. M., Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54, 480-492.

Wessely, S., A. Buchanan, A. Reed, J. Cutting, B. Everitt, P. Garety, and P. Taylor. (1993). Acting on delusions 1: Prevalence. *The British Journal of Psychiatry* 163, pp.69–76.

Wittgenstein, L. (1953). Philosophical investigations. Oxford: Blackwell.

Wittgenstein, L. (1958). *The blue and brown books*. Oxford: Basil Blackwell.

Wittgenstein, L. (1968). Wittgenstein's notes for lectures on "private experience" and "sense data." (R. Rhees, Ed.). *Philosophical Review*, 77, 271-320.

Wittgenstein, L. (1969). *On Certainty* (ed. Anscombe and von Wright). Harper Torchbooks.

Wright, S., Young, A. W. and Hellawell, D. J. (1993), Sequential Cotard and Capgras delusions*. British Journal of Clinical Psychology,* 32: 345–349.

Yarnada, K., Katsuragi, S., Fujii, I., (1999), A case study of Cotard's syndrome: stages and diagnosis, *Acta Psychiatr Scand*: 100: 396-399.

Young, A. W. (1998), *Face and Mind*. Oxford University Press.

Young, A. W. (1999), Delusions, *The Monist*, Vol. 82, pp. 571-589.

Young, A. W. (2000), The Neuropsychology of abnormal beliefs, *In Pathologies of Belief, Coltheart and Davies (Eds.), Blackwell,* pp.47-73.

Young, A. W., Reid, I., Wright, S., Hellawell, D. J. (1993). Face-processing impairments and the Capgras delusion. *The British Journal of Psychiatry*, 162(5), 695-698.

Zoumpouli, A. (2012). A contemporary approach to Jaspers' static understanding. *Dialogues in Philosophy, Mental and Neuro Sciences 5* (2):48-50.