Prospectos para uma teoria do delírio

José Eduardo Freitas Porcher

Orientado por
Paulo Francisco Estrella Faria

Tese apresentada como requisito parcial
à obtenção do grau de Doutor em Filosofia

Universidade Federal do Rio Grande do Sul
Instituto de Filosofia e Ciências Humanas
Programa de Pós-Graduação em Filosofia

Porto Alegre, Brasil

Março de 2015

Para Magda Togni

Conteúdo

Agradecimentos

Think where man's glory most begins and ends
And say my glory was I had such friends.
William Butler Yeats

Ao meu orientador, Paulo Faria. Aos queridos amigos, Pedro Pricladnitzky, Fernando Carlucci, Thiago Dihl Perin, Pakisa Togni, Natália Pasin, Marcos Chavarria, Tomás Adam e Cláudio Rabin. Aos meus pais, Paulo Porcher e Maria Teresa Freitas. Aos meus avós, Carlos e Lídia Porcher. Aos meus sogros, Sérgio e Maria Rosely Togni. À minha tia, Josemary Almeida. Aos colegas que leram e comentaram porções desta tese, Tasia Philippa Scrutton, Eduardo Vicentini de Medeiros e Pedro Pricladnitzky. Aos membros da minha banca de doutorado, Claudio Banzato, Adriano Rodrigues, César Schirmer dos Santos e Eros Carvalho. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). E, sobretudo, a Magda Togni, a quem esta tese é humildemente dedicada.

Introdução

Where there are no fixed boundaries only the timid never risk trespass.
Donald Davidson

Em 1979, um grupo de psicólogos relatou um caso incomum desenvolvido por um homem que havia sofrido uma grave lesão cerebral em um acidente de carro (Alexander et al. 1979). Após uma estada de dez meses no hospital, o paciente foi liberado para passar o fim de semana com a sua família. Depois dessa visita, o paciente passou a afirmar que agora vivia com uma "segunda" família, idêntica à sua "primeira" família, e que vivia com eles em uma casa idêntica à casa em que vivia com sua família anterior. O paciente insistiu que ambas as suas esposas possuíam o mesmo nome, a mesma aparência, o mesmo temperamento, que haviam nascido na mesma cidade e que seus irmãos possuíam os mesmos nomes. Ele descreveu sentimentos positivos com respeito a ambas as suas esposas, não demonstrando raiva ou ressentimento com respeito à deserção da sua "primeira" esposa e até mesmo expressando gratidão a esta por ter localizado uma substituta idêntica.

Esta tese versa sobre algumas das dificuldades de conceptualização do tipo de fenômeno descrito acima, denominado *delírio*.* O conceito de delírio é um dos mais importantes constructos usados para diagnosticar pacientes que, julga-se, perderam o contato com a realidade. A sua deteçcão possui importantes implicações para o diagnóstico e o tratamento de patologias mentais, bem como para a predição de comportamento e a atribuição de responsabilidade. Não obstante, o uso clínico do termo 'delírio' e a distinção entre delírios e outros estados mentais anômalos envolve diversas dificuldades. Esse fato é visível na atual edição do *Manual Diagnóstico e Estatístico de Transtornos Mentais* (DSM-5), que caracteriza o delírio como uma 'crença falsa baseada em inferência incorreta sobre a realidade externa que é firmemente mantida a

* No presente uso, 'delírio' é um termo técnico da psicopatologia , equivalente ao inglês *delusion*, ao francês *délire* e ao alemão *Wahn*, e não deve ser confundido com 'delirium'—uma síndrome neurocomportamental causada pelo comprometimento transitório da atividade cerebral em função de distúrbios sistêmicos.

despeito do que quase todos os outros creem e apesar do que constitui prova incontrovertível e óbvia do contrário. A crença não é ordinariamente aceita por outros membros da cultura ou subcultura da pessoa (i.e., não é um artigo de fé religiosa)' (American Psychiatric Association 2013, p. 819). Quase todos os aspectos dessa definição são questionáveis, como têm apontado filósofos, psicólogos e psiquiatras desde a inclusão desta na terceira edição do manual (DSM-III) em 1980. Por exemplo: uma crença verdadeira não poderia ser um delírio, contanto que o sujeito não possuísse qualquer boa razão para sustentar a crença? Delírios são necessariamente baseados em inferências? Não há delírios que não sejam sobre a realidade externa? Delírios são necessariamente mantidos com convicção? Não poderia uma crença sustentada por todos os membros de uma comunidade ainda assim ser deliróide?

Todavia, as controvérsias teóricas em torno do conceito clínico de delírio não se confinam à sua definição. Pelo contrário, as dificuldades inerentes à tarefa de definição do delírio em termos de condições necessárias e suficientes e a aparente continuidade entre o delírio e formas corriqueiras de irracionalidade sugerem que o delírio não constitui uma categoria bem delimitada de fenômenos que reflita uma distinção independente de conceptualizações e interesses humanos (ao contrário de elementos químicos, partículas subatômicas e outros exemplos paradigmáticos do que, em jargão filosófico, se denomina *espécies naturais*). Mas se a covariação das propriedades possuídas pelos membros da categoria do delírio não é constante como aquela de espécies naturais paradigmáticas, é empiricamente atestável, por outro lado, que as propriedades dos delírios covariam com *alguma* segurança e, portanto, que o delírio não se trata de uma categoria arbitrária. Há, desse modo, amplo espaço teórico para a discussão sobre a possibilidade de o delírio constituir um objeto de generalização, descoberta e explicação científica. Todavia, a respeitabilidade do delírio enquanto categoria científica é posta em cheque pela distinta possibilidade de que a atribuição do delírio tenha origem em considerações intuitivas sobre a normalidade—o que tem sido chamado de psiquiatria do senso comum (*folk psychiatry*)—e que, portanto, o delírio seja a mera formalização clínica de um conjunto de fenômenos que, ao fim e ao cabo, são dependentes do modo como nós percebemos e interpretamos certos comportamentos humanos.

Com respeito ao tipo de estado mental que caracteriza o delírio, este parece ser uma forma de *crença* anômala, como atesta a definição oferecida pelo DSM-5. Justamente por entrelaçar questões filosóficas sobre a natureza da crença e da racionalidade com a explicação de sintomas clínicos pela ciência cognitiva e pela neurobiologia, delírios têm, principalmente nas últimas duas décadas, interessado progressivamente a filósofos e cientistas da cognição. Em especial, o estatuto de crença (ou estatuto *doxástico*) dos delírios

se tornou o centro de um dos principais debates teóricos sobre o delírio, que tem como objetivo responder à questão sobre como melhor caracterizar esse estado mental. O apelo intuitivo da caracterização do delírio como crença se deve principalmente ao fato de que sua expressão linguística é normalmente condizente com a atribuição de crença ao sujeito. Por exemplo, o paciente que sofre de delírios de perseguição expressa verbalmente que está sob constante vigilância. Do mesmo modo, o paciente que sofre do delírio de Capgras expressa que seu cônjuge foi substituído por um impostor. E o paciente que sofre do delírio de De Clérambault (ou erotomania) expressa que alguma personalidade de status social elevado está secretamente apaixonada por si. Porém, diante dos colapsos flagrantes das funções cognitivas de pacientes com delírios, não é surpreendente que a implausibilidade de se lhes atribuir crenças irrestritamente tenha sido sugerida, direta ou indiretamente, desde o florescimento da nosologia psiquiátrica, com Karl Jaspers e Eugen Bleuler. Objeções ao estatuto doxástico dos delírios se apoiam sobretudo na catalogação de incongruências entre o papel funcional paradigmático da crença e aquilo que se observa em casos de delírio. Por exemplo, muitos pacientes com delírios falham em agir de forma coerente com aquilo que professam crer: o paciente que sofre do delírio de Capgras, por exemplo, raramente se preocupa com o destino do seu cônjuge abduzido, procura a polícia para registrar o seu desaparecimento, etc. Fatos como esse motivam o desenvolvimento de novas caracterizações que buscam enquadrar o delírio como outro tipo de estado mental e, assim, vão de encontro à definição do DSM.

Nos capítulos que se seguem,[†] apresentarei e analisarei os diversos problemas que concernem a conceptualização e a explicação do delírio; examinarei falhas nas principais tentativas de resposta a algumas dessas dificuldades, dando ênfase às discussões sobre a respeitabilidade científica da categoria do delírio e sobre a correta caracterização do delírio enquanto atitude proposicional; questionarei a adequação das soluções discutidas com respeito ao prospecto do desenvolvimento de uma teoria científica do delírio; e, finalmente, desenvolverei hipóteses de trabalho que visam reparar as falhas das soluções prévias, com o objetivo de oferecer suporte teórico à explicação dos fenômenos investigados.

[†] A presente tese é composta por três capítulos redigidos em inglês, que serão posteriormente submetidos à publicação como artigos separados, como faculta a Resolução no. 093/2007, da Câmara de Pós-Graduação da Universidade Federal do Rio Grande do Sul.

Capítulo 1

Conceptual and explanatory challenges of delusion

Introduction

Delusion is one of the central concepts of psychopathology, the scientific study of mental illness. It has been considered 'the basic characteristic of madness' (Jaspers 1963, p. 93), as well as the main criterion when assessing and diagnosing psychosis. The detection of delusions has profound consequences for diagnosis and treatment, as well as for the prediction of behavior and the attribution of responsibility (David 1999). Yet, for all its importance, delusion has eluded precise conceptualization. In what follows, I will explore issues concerning the nature, the explanation, and the characterization of delusion. First, the nature of delusion will be introduced through a consideration of classificatory, definitional, and ontological questions. Second, the explanation of delusion will be investigated through an examination of cognitive accounts of its etiology, and the fundamental question of whether delusions are the result of bottom-up or top-down disturbance. Third, the characterization of delusion will be discussed through an exploration of some of the difficulties inherent in framing delusional states in the language of belief. Thus, this chapter provides an introduction to the theoretical challenges involved in thinking about delusion. More importantly, however, I aim to show that 'delusion' is a highly ambiguous term, and that the phenomena to which it refers are multi-faceted. Additionally, I aim to shed light on why philosophers have taken an interest in delusions, increasingly joining the ranks of psychiatrists, psychologists, and neuroscientists in the effort to arrive at a comprehensive understanding of the phenomena.

1.1 The nature of delusion

1.1.1 The classification of delusion

Delusions occur in a variety of contexts, including paranoid schizophrenia, bipolar disorder, Alzheimer's disease, Parkinson's disease, Lewy body dementia, epilepsy, and acquired brain injury. Delusions have been grouped in many different ways.[1] The context of delusion, for example, was once a criterion for dividing delusions into organic and functional (Bortolotti 2013). A delusion was called organic if was the result of brain injury, and functional if it had no known organic cause (which usually entailed a psychodynamic or motivational explanation). The distinction is now considered to be obsolete, as the development of neuropsychiatry has increasingly lent credibility to the view that all delusions have an organic basis, even though some have not been precisely identified yet.

Delusions are perhaps most intuitively classified according to their content—that is, according to what the delusion is about. Not only pre-twentieth-century inventories bear witness to this characteristic (Berrios 1996), but it also has made its way into current classifications. For example, the section 'Schizophrenia Spectrum and Other Psychotic Disorders' of the current edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) states that the content of schizophrenic delusions may include a variety of themes, such as persecutory, referential, grandiose, erotomanic, nihilistic, and somatic. Persecutory delusions involve the conviction that one is being, or is going to be, harmed or harassed by an individual or organization; delusions of reference involve the conviction that certain gestures, comments and environmental cues are directed at oneself; grandiose delusions involve the conviction that one has exceptional abilities, wealth, or fame; erotomanic delusions involve the conviction that another person, usually of high status or famous, is in love with the patient; nihilistic delusions involve the conviction that a major catastrophe will occur; and somatic delusions focus on preoccupations regarding health and organ function (American Psychiatric Association 2013, p. 87). The thematic families listed in the DSM are some of the most clinically common—especially persecutory delusions and delusions of reference—but the list is not meant to be exhaustive. Indeed, it only scratches the surface of the thematic variety of delusion.

[1] In the interest of conciseness, I will present only five ways of classifying delusions. There are, however, many other classificatory distinctions available, such as mood-congruent vs. mood-incongruent (Kumazaki 2011), authored vs. unauthored (Bortolotti and Broome 2008) and individual vs. group delusions (Shimizu et al. 2007). For a comprehensive introduction to the varieties of delusion, cf. Radden (2011, pp. 17-39).

In his lauded *General Psychopathology*, Karl Jaspers effected a shift in the classification of delusions from their content to their formal or structural features, such as their comprehensibility. For Jaspers, the psychiatrist's inability to achieve an empathetic understanding of the patient's experience was the true sign of madness and it was the chief criterion for his distinction between primary delusions (or delusions proper) and secondary delusions (or delusion-like ideas). Jaspers maintained that the former cannot be understood phenomenologically and originate in what he describes as a 'transformation in our total awareness of reality' (1963, p. 95) while the latter originate in understandable ways from experience.

This shift from an extensional to an intensional classification is felt in the distinction between bizarre and nonbizarre delusions—a distinction of some clinical importance, as the DSM treats the presence of bizarre delusions as the heaviest-weighted clinical criterion of schizophrenia.[2] According to the DSM, delusions are deemed bizarre when two conditions are met: first, they are clearly implausible and incomprehensible to same-culture peers; second, they are not derived from ordinary life experiences (American Psychiatric Association 2013, p. 87). Instances of delusion that seem to satisfy these criteria abound in the clinical literature. For example, one patient had the delusion that there was a nuclear power station inside his body (David 1990); another, that he was both in Boston and in Paris at the same time (Weinstein and Kahn 1955). Much more common, however, are delusions that do not satisfy the criteria for bizarre delusion; that is, delusions that appear understandable and derived from ordinary life experiences. As an example, the DSM alludes to the conviction that one is under surveillance by the police, despite a lack of convincing evidence.

Finally, a recent and useful distinction divides the set of delusions into monothematic and polythematic (Davies et al. 2001). A monothematic delusion is one that is specific to a particular theme. It contrasts with polythematic delusion, in which case patients exhibit many delusions concerning a variety of themes. Monothematic delusions are typically not elaborated and not integrated (or not completely integrated) with the rest of the patient's beliefs, while polythematic delusions are both elaborated and integrated. Monothematic delusions are commonly the consequence of acquired brain injury. Examples of delusions that present as monothematic include those that are referred to as Delusional Misidentification Syndromes (Christodoulou

[2] Bell and colleagues (2006) reviewed the inter-rater reliability of the category of bizarre delusions, concluding that it was inferior to that for delusions "in general" and that the concept was inadequate for scientific usage. Cermolacce and colleagues (2010) point out, however, that only a small fraction of schizophrenia patients receive their diagnosis because of the presence of bizarre delusions (4%–8%).

1986), such as Capgras delusion, Fregoli delusion, and reduplicative paramnesia.[3] Polythematic delusions are often and appropriately referred to as delusional systems, being most commonly associated with schizophrenia (Coltheart 2013).

Capgras delusion, described by Joseph Capgras and Jean Reboul-Lachaux in 1923, involves the conviction that one's loved ones (typically one's relatives or spouse) have been replaced by doubles—impostors which are usually human, but in some cases may be ghosts, aliens, or robots (Rodrigues et al. 2013).[4] Fregoli delusion, described by Paul Courbon and Gustave Fail in 1927, typically involves the conviction that strangers are actually familiar individuals in disguise, or that different people are in fact a single person who changes appearance or is in disguise (Mojtabai 1994).[5] Finally, reduplicative paramnesia, named by Arnold Pick in 1903 and, in all indication, first described by Charles Bonnet in 1788 (Förstl and Beats 1992), typically involves the conviction that a location has been duplicated, existing in two or more places simultaneously, or that it has been relocated to another site.

Perhaps the most famous case of polythematic delusion in psychiatric history remains that of Daniel Paul Schreber, an appellate judge in the kingdom of Saxony who spent thirteen years in mental asylums and wrote of his experiences with schizophrenia in *Memoirs of My Nervous Illness* (Schreber 1903)—a fame that was due in no small part to the fact that his account was the subject of a major study by Sigmund Freud (1911), as well as being extensively explored by Eugen Bleuler (1912), and offered as an example of schizophrenic incomprehensibility by Jaspers (1913).[6] The core of Schreber's delusional system included the conviction that he had a mission to redeem

[3] Providing an exhaustive list of the delusions that present as monothematic is a difficult task, not only because of their sheer multitude, but because some of them, such as the Reverse Othello Syndrome (Butler 2000), have been reported but once. Nevertheless, the list includes the other delusions commonly grouped under Delusional Misidentification Syndromes—intermetamorphosis (De Pauw and Szulecka 1988), the delusion of subjective doubles (Christodoulou 1978) and mirrored-self misidentification (Coltheart 2011)—as well as Cotard delusion (Young and Leafhead 1996), erotomania (Berrios and Kennedy 2002), and the delusions of alien control and of thought insertion (Frith 1992). [4] Though the visual misidentification typical of Capgras delusion has also been reported with regard to animals (Somerfield 1999) and even objects (Abed and Fewtrell 1990). There also have been cases of delusional voice misidentification, a disorder referred to as "blind Capgras" (Reid et al. 1993; Dalgalarrondo et al. 2002) [5] Courbon and Fail named it after Italian actor Leopoldo Fregoli, a protean actor who, according to Matthew Solomon, 'was contrasted with other quick-change performers because his virtuoso talents of impersonation set him apart from others who merely made costume changes with dexterity' (2000, p. 7). [6] Other notable first-person accounts of systematic delusion include *Narrative on the Treatment Experienced by a Gentleman during a State of Mental Derrangement* (Perceval 1840), *Autobiography of a Schizophrenic Girl* (Sechehaye 1951), and *The Diary of Vaslav Nijinsky* (Nijinsky 1995).

the world and to restore mankind to their lost state of bliss. In order for this to happen, he insisted, divine forces were preparing him for a sexual union with God by changing him into a woman, so he could give birth to a new race of humanity. Schreber never disavowed what he termed 'my so-called delusions' and died in an asylum in 1911 (Sass 1994).

1.1.2 The definition of delusion

To provide a definition of delusion that satisfies the needs of both psychopathological theory and clinical practice is a difficult task. The first two editions of the DSM—DSM-I (1952) and DSM-II (1968)—did not provide one, but with the inclusion of the section 'Glossary of Technical Terms' in the DSM-III (1980), the manual came to define delusion as follows:[7]

> A false belief based on incorrect inference about external reality that is firmly held despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not ordinarily accepted by other members of the person's culture or subculture (i.e., it is not an article of religious faith). When a false belief involves a value judgment, it is regarded as a delusion only when the judgment is so extreme as to defy credibility. (American Psychiatric Association 2013, p. 819)

Reflection upon and attention to the clinical literature raise a number of difficulties concerning this attempt at a definition (Garety and Hemsley 1997; Spitzer 1990; Leeser and O'Donohue 1999). Does delusion have to be false?[8] Consider a case of Othello syndrome—the delusion that one's spouse or sexual partner is being unfaithful—discussed by Jaspers (1913), in which the stress provoked by living through the morbid jealousy of her husband causes the patient's wife to find consolation in another man's arms, thereby verifying the patient's delusion. Nothing in the patient's mind has changed: he still holds that his wife is unfaithful without having any evidential justification. So it is not the truth-value of the proposition or propositions held by the delusional that is epistemologically interesting to the characterization of delusions, but the fact that they are 'sustained despite what constitutes incontrovertible and

[7] The citation is from the current edition of the DSM, DSM-5, but the definition has remained practically unchanged. The only revision since the DSM-III was the suppression of the qualification 'personal belief' in the DSM-IV-TR (2000). [8] One may also ask: is a delusion always a belief? This question will be taken up in section 3 of this chapter and in more detail in chapter 3 of this dissertation.

obvious proof or evidence to the contrary', etc.[9] As Golda Meir is reputed to have quipped after being accused of being paranoid by Henry Kissinger for hesitating to grant further concessions to the Arabs during the 1973 Sinai talks, 'Even paranoids have enemies' (Berke et al. 1998, p. 1).

Does delusion have to be based on inference?[10] As Martin Davies and colleagues (2001, p. 134) observe, a subject might form a delusional belief simply by taking an anomalous perceptual experience to be true, and it is not obvious why this might involve an inferential step. Furthermore, Philip Gerrans has advanced a theory that relieves the emphasis on hypothesis confirmation to which the inferential view alludes, proposing that processes of selective attention and recall exert their effects instead on autobiographical narrative. In his words, 'Someone with a delusion is not a mad scientist but an unreliable narrator' (2009, p. 152). Therefore, the inferential nature of delusion formation is a point of contention. This raises the further question of whether definitions of mental disorders should include explicitly theoretical elements.

Does delusion have to be about external reality? Consider delusions that concern the subject's own body—such as manifestations of Cotard's syndrome in which the patient affirms that some of her internal organs are missing (Berrios and Luque 1995), or somatoparaphrenia, which involves the denial of ownership of one or more of one's limbs or sometimes an entire side of one's body (Vallar and Ronchi 2009)—or delusions that concern the subject's own thoughts—such as thought insertion, in which the subject reports that another's thoughts occur in her own mind without her volition (Fulford 1993). Whether it is about "external" or "internal" reality—a terminology so vague as to merit scientific disrepute—is of no consequence to the delusional character of a belief.

Does delusion have to be firmly sustained? While that may be the case in many if not most manifestations, the conviction of delusional subjects is subject to fluctuation. Davies and colleagues (2001) observe that at least some delusional patients show appreciation of the implausibility of their delusional beliefs, quoting an excerpt of an interview with a patient with reduplicative paramnesia who thought that his house and family had been replaced by duplicates:

> E: Isn't that [two families] unusual?
> S: It was unbelievable!

[9] In addition, the propositions that some patients appear to believe may be utterly unfalsifiable, such as John Nash's assertion that he was the left foot of God and that God was walking on the earth (Nasar 1998, p. 258). [10] One may also ask: does the inference have to be incorrect? This question will be taken up in section 2.1 of this chapter.

E: How do you account for it?
S: I don't know. I try to understand it myself, and it was virtually impossible.
E: What if I told you I don't believe it?
S: That's perfectly understandable. In fact, when I tell the story, I feel that I'm concocting a story . . . . It's not quite right. Something is wrong.
E: If someone told you the story, what would you think?
S: I would find it extremely hard to believe. I should be defending myself. (Alexander, Stuss and Benson 1979, p. 335)

This kind of explicit ambivalence on the part of the patient is starkly manifested in the testimony of John Custance, a former Royal Navy intelligence officer who suffered from bipolar disorder and wrote of his experiences with mental illness in *Wisdom, Madness and Folly: The Philosophy of a Lunatic*: 'Of course it is all ... pure imagination ... I know perfectly well that in fact I have no power, that I am of no particular importance and have made rather a mess of my life. ... Moreover, psychologically speaking, I know that my delusions of grandeur are merely compensations for the failures and frustrations of my real life' (1952, p. 52).

Does delusion have to contradict what almost everyone else believes? Or: does the attribution of delusion have to take into consideration the person's culture or subculture? Davies and colleagues object: 'If a bizarrely implausible belief is formed and sustained in ways that are characteristic of delusions, then it seems that, for the purposes of psychological theory, it should be grouped together with delusions even if many other subjects believe the same thing' (2001, p. 133). However, as *ad hoc* a clause as it may seem, cultural exemption may make sense of the fact that we do not think that individuals who belong to cultures such as that of the Uduk people are delusional. Dominic Murphy reports the fieldwork done by Wendy James (1988) in the Sudan, where it is believed that trees convey information: 'You can learn what they know by burning an ebony twig, dipping it in water and reading the pattern of ashes in the water' (2013, p. 119). The cultural exemption clause encodes into the definition of delusion the fact that we would attribute a delusion to someone in our culture if they held that they gathered knowledge about the plans of witches from trees, but not with respect to the Uduk. However, as with the inferential nature of delusion formation, the cultural exemption clause is again a point of contention.

Does delusion have to occur in the face of incontrovertible and obvious proof or evidence to the contrary? Consider the case of mirrored-self misidentification—the delusion that one's reflection in the mirror is not one's own

(Coltheart 2011). It sometimes is accompanied by the conviction that whoever the person in the mirror is, he or she is following the subject around. Now, are these patients in possession of 'incontrovertible and obvious proof or evidence' that, although they fail to identify the face in the mirror, it is nevertheless theirs?[11] Consider that just as not all hallucinatory symptoms lead to delusion, an otherwise normal subject presented with the anomalous experience of not recognizing oneself in the mirror would not arrive at the belief that, say—although the mirrored person is waving just like I am, wearing the same clothes, sporting the same hairstyle, etc.—that person is not me. In addition to these overriding facts (which point to the great plausibility that there is something wrong with *me*), the testimony of each and everyone of one's epistemic peers would also weigh in heavily in the reasoning of a person whose thoughts did not mark the presence of some deficit, or bias, or both.[12] So imperviousness to evidence does indeed seem to be a central feature of delusion.

Indeed, delusion is often not only impervious to evidence that tells against it, but it also persists in spite of bad consequences—even self-perceived harmful and imprudent consequences (Mojtabai and Nicholson 1995). A final observation of the inadequacy of the DSM definition is that it captures exclusively epistemological features, failing to take the disruption of day-to-day functioning into account (McKay et al. 2009)—that which is typically the focus of clinical concern and treatment. It ultimately ignores the fact that, as George Graham sums up, 'Living through a delusion *hurts* a person' (2010, p. 203, my emphasis).

1.1.3 The ontology of delusion

The fact that the standard definition of delusion has proved so problematic raises the question of whether delusion can ever be given a proper definition in terms of necessary and sufficient conditions. Put another way, it raises the question of whether all the various types of delusion we have discussed share a common essence, something to which we could refer in order to ultimately decide if something is or is not a delusion. Is delusion a class of things akin to quarks, noble gases, and tigers, in their suitability for the purposes of scientific investigation? Does delusion as a kind "carve nature at its joints," latching on to a real distinction in nature? In other words, is delusion a

[11] A similar question could be raised concerning other delusional misidentification syndromes such as Capgras. [12] The question of how to explain the genesis of delusion will be taken up in section 2 of this chapter.

natural kind?[13]

'Natural kind' is philosophical jargon and, therefore, the question 'Is delusion a natural kind?'[14] is a loaded one. Beyond depending on an investigation of the characteristics of delusions as a whole, an answer to it will be determined by one's view of what requisites a class of things should fulfill in order for it to be considered a natural kind. The traditional account of natural kinds is represented by various forms of *essentialism*, which usually involves three main tenets (Ereshefsky 2009). First, all and only the members of a kind share a common essence. Second, that essence is a property, or a set of properties, that all the members of a kind must have. And third, a kind's essence causes the other properties associated with that kind. So, for example, the essence of gold is gold's atomic structure, and that atomic structure occurs in all and only pieces of gold. That structure is a property that all gold must have as opposed to such accidental properties as being valuable to humans. And the atomic structure of gold causes pieces of gold to have the properties associated with that kind, such as readily dissolving in mercury at room temperature, conducting heat and electricity, and being unaffected by air and moisture.

As essentialism holds that natural kinds exist independently of our classifications, it behooves scientists to discover their inherent essences and classify them accordingly. The conceptualization of scientific kinds as essentialistic natural kinds has indeed been applied with success, especially in physics and chemistry, but is it applicable to psychiatric kinds, or even biological kinds?[15] Can psychiatric disorders and symptoms be exhaustively defined by fixed and inherent properties? Can delusion, in light of the fact that the conditions in its standard definition are not necessary or even jointly sufficient?

On the other hand, assuming that there is no essential criterion or set of criteria for being a delusion does not, by itself, entail that delusion as

[13] The issues of the natural kind status of delusion and the prospects for a scientific theory of delusion are taken up in more detail in chapter 2 of this dissertation. [14] One may also ask whether or not some *subtypes* of delusions are natural kinds. Here, however, I will concern myself with introducing the more general question of whether the whole category of delusions as such constitutes a natural kind. If it does, then it will be a generic kind with more specific natural kinds in its extension, like metal and magnesium, respectively (Samuels 2009, p. 76). [15] Consider biological species, the main candidates for natural kindhood in biology. While the existence of various evolutionary forces does not rule out the possibility of a trait occurring in all and only the members of a species, it is extremely unlikely that biological species have essences (Ereshefsky 2009). Three main views have been advanced in response to this observation: denying that species are natural kinds and looking elsewhere in biology for kinds with essences (Hull 1978); arguing that species are indeed kinds with essences, but that their essences are of a nontraditional variety (Okasha 2002); and, as we will see below, arguing that natural kinds do not require the sort of essences implied by essentialism (Boyd 1999).

a kind is nothing but an arbitrary clustering of properties. 'Delusion' picks out reasonably stable, nonarbitrary patterns, and application of delusion as a classification seems justified by its usefulness for clinical purposes (Bell et al. 2006). In keeping with these observations, Peter Zachar (2000) proposes that mental disorders be conceptualized as *practical kinds*. As an example, Zachar (2014b, pp. 154-5) alludes to the distinction between an adult and a child. Although the kinds 'adult' and 'child' are not in themselves sharply demarcated, the uses for which we deploy them will determine where their boundaries should be drawn. Consequently, many distinctions between adults and children are context-dependent. For example, if our aim is to decide who is able to vote, engage in consensual sex, get married, be sent to prison, drink alcohol, or enter into a legal contract, each of those considerations will result in different ways of demarcating adulthood (Horwitz and Wakefield 2012, p. 53).

Is Zachar right in arguing that psychiatric kinds are practical kinds that pick out mind-dependent distinctions? Or do they pick out mind-independent distinctions in nature? Importantly, what is the relevant sense of mind-independence with regard to the characterization of natural kinds? Richard Samuels argues that it is what Sam Page (2006) calls *individuative independence*: 'Roughly put, a kind, K, is individuatively independent if it is circumscribed by boundaries that are totally independent of where we draw the lines. In other words, individuatively independent kinds are the sorts of kinds whose existence does not (metaphysically) depend on how we categorize things' (2009, p. 54). Page illustrates his concept by alluding to the individuation of the night sky into constellations: 'Though it is *prima facie* plausible that reality is individuated intrinsically into stars, reality is not individuated intrinsically into constellations, since it is people who divide the night sky into constellations' (2006, p. 328). Furthermore, although the International Astronomical Union divides the celestial sphere into 88 official constellations, there can be as many different star maps as there are people willing to point out a few stars and give the cluster a name.

With respect to individuative independence, then, Zachar's practical kinds model has the import of making psychiatric kinds out to be akin to constellations rather than stars. However, since psychiatric kinds are manifold and differ greatly with respect to validity, it is possible for some to be mind-dependent kinds, and for others to turn out to be mind-independent—and, among those that are merely mind-dependent kinds, some may be practical kinds in Zachar's sense, while others may not even rise to such a status. With regard to the specific case of delusion, three considerations put pressure on

the assumption that it constitutes a mind-independent kind.[16] First, delusions may be an artifact of our folk psychology, our commonsense mode of thought about mental states and processes, as Murphy proposes:

> Whether or not something is a delusion is a matter of how it strikes us, and that depends on how well it comports with our understanding of what people are like, both in general terms and within our culture. It does not depend on some psychological mechanism or a formal property of beliefs. (2006, p. 180).

Murphy's observation that being a delusion is a response-dependent property stems from reflection on the attribution of delusion. He argues that a delusion is attributed to a subject when our explanatory resources run out and we cannot make sense of how and why someone has a certain belief: 'a delusion is a belief that is acquired in ways that defeat our expectations about belief acquisition' (2013, p. 117).

Second, as I have pointed out when discussing the cultural exemption clause in the DSM definition of delusion, what is considered a delusion in one place (or at one time) may not be considered in another. This ties neatly with Murphy's theory of delusion attribution as a failure of folk epistemology to account for someone's acquiring a belief, as what will count as a reason for holding a belief will ultimately depend on the context of attribution. Consider again the example of Sudan's Uduk-speaking peoples. Believing that ebony trees can eavesdrop on conversations and that information about such conversations can be read off from them through divination will count as a reason for refusing to conduct a conversation near an ebony tree (Boyer 2001, p. 69). In Uduk society, in contrast with Western society, this kind of reasoning will be understandable. To the extent that what is a delusion depends on what beliefs are socially prevalent in the context of attribution, cultural relativity suggests that being a delusion is a response-dependent property.

Third, delusions are normatively assessable: to be deluded usually (if not necessarily) means that something is *wrong*. While this does not necessarily entail mind-dependence, if the norms to which the assessment of delusion is subject are in any way social, then the very existence of delusions would turn out to depend on our cultural modes of thought. In other words, the boundaries of delusion would be at least partly dependent on where we draw the lines. Hence, delusion would not be an individuatively independent kind. But are the norms that govern delusion social?

[16] These will be discussed at greater length in section 3 of chapter 2 of this dissertation.

Delusions may be subject to at least two kinds of norms, namely, medical norms and norms of rationality (Samuels 2009). On the one hand, it is difficult not to accept that delusions are typically, if not always symptomatic of pathology—and even the least socially laden theories of mental disorder accept that the notion of harm should be understood in sociocultural terms (Wakefield 1992). On the other hand, it is hard to avoid the conclusion that some, if not all, delusions are epistemically irrational[17]—although whether norms of rationality are even partially socially constructed is much more controversial.[18]

Against these threats, Samuels has argued that the line of reasoning present in the mind-dependence objections to the natural kinds status of delusion conflates the metaphysics of delusion with its epistemology:

> The relevant metaphysical issue concerns the *nature* of delusions: roughly, what is it to be a delusion. The relevant epistemic question concerns the *evidential basis* for our judgements about delusion: roughly, the sorts of evidence we invoke in judging that someone is deluded. (2009, p. 68, my emphases)

However, even if such evidential basis were necessarily linked to culture-bound folk epistemologies and mind-dependent norms, he argues, there remains the modal point that this alone would not establish a necessary link between what it is to be a delusion and our judgments about what it is to be a delusion—the connection may be a contingent one.

Ultimately, the importance of investigating what kind of thing delusions are lies in determining if they constitute an appropriate category for the purposes of scientific inquiry, such as inductive generalization, empirical discovery, and mechanistic explanation. Toward that end, the essentialist demand that all and only members of a kind share intrinsic properties as a matter of metaphysical necessity may be overly restrictive, since many kinds that successfully figure in scientific practice, such as biological taxa, do not meet these conditions. Partly for this reason, the predominant opinion in philosophy of science is that such a *sortal* notion of essence should be replaced by a merely *causal* notion that entails only the existence of a set of empirically discoverable causal mechanisms that explains the covariation of the charac-

[17] The reason I qualify 'irrational' here is because the aspects we have discussed so far relative to the DSM definition are decidedly epistemic (as opposed to *procedural* and *agential*, for example): delusions seem to lack evidential support, fly in the face of strong counterevidence, etc. For an examination of delusions with respect to epistemic, procedural, and agential rationality, see Bortolotti (2010). [18] Indeed, most philosophical accounts of rationality do not reduce rational norms to social norms (e.g. Howson and Urbach 1993, Nozick 1993, Stich 1990).

teristics or symptoms co-instantiated by instances of a kind (Samuels and Ferreira 2010).[19]

Settling the dispute about whether delusion constitutes a practical kind or a natural kind in the liberal sense will depend, then, on ascertaining through exploratory research whether delusion as a kind is individuated by a causal essence. A strong indication that this is the case would be for explanations of delusions to exhibit some kind of unity. So far, such unity remains a distant goal and the options are all still on the table, including the possibility that delusion as a generic kind picks out a merely practical distinction while some of its subtypes possess the individuative independence and causal unity required of natural kinds.

However, even if the investigation of the neurobiological causes of delusion reveals that delusion as a such is not nondisjunctively characterizable in the vocabulary of biological neuroscience, explanatory unity may be found at other levels of explanation. As we will see in the next section, causal explanations of delusion have mostly focused on computational processes at the cognitive level. Ultimately, however, given that 'many factors are implicated in delusion development, and the contribution of each in individual cases varies' (Freeman and Garety 2006, p. 207), seeking an explanation that *integrates* the various levels of description—from neurobiological to phenomenological—may turn out to be our best chance to arrive at a unified theory of delusions (Gerrans 2014).

1.2 The explanation of delusion

There is no generally accepted theory of the etiology of delusions. The construction of an explanatory model of acquisition remains one of the main research controversies surrounding delusion. Attempts to provide a cognitive explanation by and large assume that delusions are beliefs formed in response to perceptual or sensory experiences. Single-factor accounts make delusional subjects out to be broadly instrumentally rational, attempting to explain delusions as normal responses to abnormal experiences. Two-factor accounts are based on the conviction that we need to postulate a second factor to explain why delusions are maintained in the face of extraordinary implausibility, strong counterevidence, and the testimony of one's peers. Finally, it has been also a matter of dispute whether delusions always involve bottom-up causation or if at least some delusions are better understood as a result of top-down disturbance.

[19] This is best exemplified by the most influential and widely adhered to theory of natural kinds, the *homeostatic property cluster* theory (Boyd 1991).

1.2.1 Maher's one-factor account

Brendan Maher has offered a model of delusion formation that emphasizes the role of anomalous experience. He summarizes his account in the following propositions:

1. Delusional thinking is not in itself cognitively aberrant. This means that the cognitive processes by which delusions are formed are in no important respect different from those by which normal beliefs are formed. Parenthetically, in neither case are beliefs typically formed by a process of syllogistic deductive reasoning.

2. Delusions are like scientific theories to the extent that they serve the purpose of providing order and meaning for empirical observations.

3. As in the case of normal scientific theorizing, the necessity for a theory arises whenever nature presents us with a puzzle. Puzzles arise when predictable events fail to occur and/or unpredicted events do so in their place, i.e., when observation is discrepant with expectation. (2001, p. 321)

Because delusion formation is not significantly different from the process of forming normal beliefs, Maher maintains that if delusions are pathologies of belief, then the locus of pathology lies in experience and not in the subject's reasoning processes (Maher 1999). When the anomalous experience occurs, it attracts the subject's attention and gives rise to an experienced feeling of significance accompanied by some tension. This tension motivates a search for explanation which is continued until some explanation has been found. While the explanation may be less than fully adequate, it will reduce anxiety and bring relief, inasmuch as a partially defective or incomplete explanation is experienced as better than no explanation at all. Furthermore, not only may delusional hypothesis formation be likened to a scientist's hypothesis formation, but resistance to let go of the explanation (i.e. the delusion) on the part of the delusional subject may be likened to a scientist's resistance to the disconfirmation of her theories. Hence, delusions are perfectly normal

responses and the delusional subject is understood as broadly rational.[20]

Maher (2001) presents three main sources of evidence for his account. First, he notes that delusions occur in an wide array of medical and psychological conditions in which the patient has no prior history of cognitive impairment (Manschreck 1979).[21] Second, he cites evidence that delusions can be induced in normal subjects under anomalous environmental conditions.[22] Finally, Maher's model has provided a framework for the cognitive therapy of delusions that has been effective in some cases (Chadwick and Lowe 1990). Nevertheless, the evidence invoked by Maher pales in comparison to the explanatory problems faced by his account.

The first problem for the one-factor account has to do with the insufficiency of abnormal experiences to explain the formation of delusion, since abnormal experiences do not always elicit delusion.[23] On the one hand, hallucinatory experiences do not necessarily evoke delusional interpretation—which suggests the involvement of other factors in delusion formation (Krabbendam et al. 2005, p. 184). On the other hand, there is evidence that suggests that there are subjects who suffer from the same type of brain damage, and plausibly have the same experiences, as the subjects who develop certain monothematic delusions, but who do not form or at least do not accept delusional explanations to account for their experience. As Davies and colleagues have observed:

> On Maher's view ... [i]t follows that anyone who has suffered neuropsychological damage that reduces the affective response to faces should exhibit the Capgras delusion; anyone with a right

[20] Note, however, that commitment to a one-factor account does not need imply commitment to the thesis that delusion acquisition is rational. Gerrans has argued that one-factor accounts should not be thought of as claiming that a delusional subject is rational in the sense of conforming to idealized norms of deductive or probabilistic reasoning: 'Clearly, it is irrational, measured against canons of inferential consistency, to believe a proposition for which you have conclusive falsifying evidence (for example, to believe that you are dead, as Cotard patients often claim). Rather, the one-stage theorist should be understood as claiming that the actual psychology of belief formation, which departs considerably from ideal rationality, functions in the same way in normal and delusional subjects' (2002, p. 48). [21] This is most often the case for delusional patients with acquired brain injury (both traumatic and nontraumatic). For example, a recent case study reported that a 76-year-old male patient without prior history of cognitive impairment was admitted to the stroke unit of a Lisbon hospital with a left occipital hematoma and presented with persecutory delusions in the second day of hospitalization (Frade et al. 2013). [22] For example, elderly subjects made partially deaf by hypnotic suggestion, but kept unaware of the source of their deafness, became more paranoid as indicated by their scoring higher than controls on a variety of assessment measures such as the Minnesota Multiphasic Personality Inventory (Zimbardo, Andersen and Kabat 1981). [23] Furthermore, some delusions are thought to occur in the absence of any anomalous experiences (Chapman and Chapman 1988).

> hemisphere lesion that paralyzes the left limbs and leaves the subject with a sense that the limbs are alien should deny ownership of the limbs; anyone with a loss of the ability to interact fluently with mirrors should exhibit mirrored-self misidentification, and so on. However, these predictions from Maher's theory are clearly falsified by examples from the neuropsychological literature.[24] (Davies et al. 2001, p. 144)

The second problem has to do with the insufficiency of abnormal experiences to explain the maintenance of delusion—a crucial feature of any appropriate model of delusion, as Max Coltheart has pointed out (2007, p. 1044). Even if an anomalous experience provides an answer to the question 'where did the delusion come from?', it does not provide any answer to the question 'why does the patient not reject the belief?'. So for the maintenance, as well as the formation of delusion to be accounted for, it seems that a second, nonexperiential factor must be postulated.

1.2.2 Multi-factor accounts

The term 'cognitive neuropsychiatry' was first used in 1991 to refer to the application of the methods of cognitive neuropsychology to psychiatric disorders (Aimola Davies and Davies 2009). In 1996, the journal *Cognitive Neuropsychiatry* was launched, with the journal editors noting some of the changes of approach that needed to be attended to with the shift from cognitive neuropsychology: 'We need to think of excesses as well as deficits; transient rather than stable phenomena; distortions and biases rather than striking quantitative or apparent qualitative differences' (David and Halligan 1996, p. 2).

The poster-child for the approach of cognitive neuropsychiatry is Hadyn Ellis's and Andrew Young's account of delusional misidentification. As we saw earlier, in Capgras delusion individuals present with the conviction that someone close to them has been replaced by an identical impostor. Ellis and Young (1990) hypothesized that the Capgras delusion results from damage to a neurological system involved in orienting responses to seen faces based on their personal significance. Their explanation is based in a two-route model of face processing, involving both a visuo-semantic pathway that processes semantic information about facial features, and a visuoaffective pathway that produces a specific affective response to familiar faces. So, in this model,

[24] With regard to Capgras delusion, cf. Tranel and colleagues (1995). With regard to denial of ownership of limbs, cf. Bisiach and Geminiani (1991, p. 20). And with regard to mirrored-self misidentification, cf. Breen and colleagues (2000, pp. 87, 91-92, 101-102).

face recognition functions like a logical AND gate, requiring two sorts of input (Murphy 2006, pp. 172-173). The semantic input is the one missing in prosopagnosia, a perceptual disorder in which individuals are unable to recognize familiar faces and, in some cases, their own face in the mirror. The affective input is the one missing in Capgras delusion and, hence, Capgras results from the patient's (subpersonal) attempt to reconcile the fact that, for example, the person standing in front of him looks exactly like his wife, with the utter absence of an emotional response toward his wife's face.[25]

So Ellis's and Young's model raises the same kinds of question raised by Maher's account, namely, why do Capgras patients but not patients with prosopagnosia come up with a delusional explanation for their abnormal experience, and why do they hold on to it? In response to the need for nonexperiential factors to account for acquisition and maintenance, Tony Stone and Andrew Young (1997) have proposed a two-factor explanation of both Capgras and Cotard delusion in which the second factor consists in a reasoning *bias*. Stone and Young hypothesized that Capgras and Cotard were two ways of responding to the same experiential anomaly, namely, a feeling of unfamiliarity or strangeness. In their model, the fundamental difference between Capgras and Cotard patients has to do with which kind of attributional style an individual is prone to adopt. If it is an *externalizing* bias, then the subject ascribes the blame for the feeling of unfamiliarity to external factors, thus adopting the Capgras explanation according to which loved ones have been replaced by impostors. If it is an *internalizing* bias, then the subject ascribes the blame to internal factors, thus adopting the Cotard explanation according to which he or she is dead or somehow unreal. Therefore, Stone's and Young's two-factor account has the added advantage of explaining why these delusions have the content that they do.[26]

But are reasoning biases sufficient to account for the maintenance of delusions? By relying on normal cognitive biases, the bias model seems to incur the same explanatory insufficiency of Maher's one-stage model by making the delusional subject out to be broadly rational and failing to account for the persistence of delusions in the face of counterevidence (Bermúdez 2001). In other words, the question remains as to why delusional patients do not revise their beliefs. In keeping with this, Robyn Langdon and Max Coltheart (2000) argue that no reasoning style can be as pathologically immune to counterevidence as delusions are and that, consequently, the etiology of delusions must

[25] Ellis's and Young's hypothesis was tested by Ellis and colleagues (1997) measuring skin-conductance response in five patients, confirming that the affective response was indeed absent. [26] However, Stone's and Young's model holds more promise as an account of Capgras than as a double account of Capgras and Cotard, since, as Gerrans (2000, p. 112) notes, Cotard seems to involve a global (rather than focal) alteration of affective experience.

include a reasoning *deficit* in addition to abnormal experiences and reasoning biases—that is, delusions arise when the normal cognitive system which people use to generate, evaluate, and adopt beliefs is damaged. In a similar vein, Davies and Coltheart (2000) hypothesize that delusion is caused by retention at all costs of an externalizing attributional hypothesis which is sustained (rather then rejected) by a deficit in belief revision.

In what does this deficit exactly consist, and how much damage is necessary for the persistence of delusion? Manifestly, the breakdown of the delusional subject's belief revision capacities is partial, since even in the most dramatic psychotic cases (such as that of Schreber), there is some preservation of normal, alongside delusional reasoning. Murphy (2006, pp. 176-180) observes that deficit models fail to actually specify deficits in terms that go beyond folk psychology and make contact with a theory of central systems, and forcefully argues that these cognitive models reify commonsense capacities and assume that systems exist to underwrite them without providing independent justification for the postulation of such systems:

> Langdon and Coltheart, for instance, argue that normal belief revision depends on two types of sensory information; some we attend to because of "heightened personal salience," and some we are automatically "oriented towards because it is discordant with our prior experience of how the world should be." But this is just the view that people normally change their mind when they learn something important or surprising, which we already knew. The psychological clout comes when Langdon and Coltheart suggest that "two distinct mechanisms" exist to carry out separate monitoring tasks that correspond to these two psychological traits. But these putative mechanisms ... [are] no more than names for some aspect of reasoning that a theorist has chosen to identify. (2006, p. 177, references omitted)

Finally, all models discussed so far assume that all delusions can be understood as explaining away anomalous experiences. This is referred to as *empiricism* (or the *bottom-up* approach) toward the etiology of delusion. However, whether delusions are always grounded in anomalous experiences is a matter of dispute. While delusions like thought insertion, the delusional misidentification syndromes, and some forms of anosognosia are plausibly prompted by anomalous experiences, there are various delusions that resist such a treatment. For example, consider the case of a patient who noticed three marble tables in a café and was suddenly convinced that the world was coming to an end (Sass 1992, p. 153). As John Campbell (2001, p. 95) notes, it seems impossible to understand how any experience at all, still less an

experience of marble tables, could be explained by, or aid in the verification of, the proposition 'The world is ending.'

With regard to monothematic delusions, José Bermúdez (2001, p. 473) raises the same problem for persecutory delusions, observing that although there is a precedent for analyzing them in part as mechanisms of self-defense responding to abnormally low self-perception (Bentall, Kaney and Dewey 1991), there is no reason to think of such self-perception as an anomalous perceptual experience. And Murphy (2006, p. 174) points to erotomania, also known as de Clérambault's syndrome, which involves the conviction that one is loved by someone of high status, often an inaccessible figure. In the original case described in 1920 by Gaëtan Gatian de Clérambault, a French woman became convinced that George V reciprocated her love. She would make several trips to London, standing outside the gates of Buckingham Palace and interpreting such things as curtain movements as signals from the king. Even more so than in the case of persecutory delusions, there does not seem to be any reason to assume that erotomania explains away anomalous experiences.

1.2.3 The direction of causal explanation

In contrast with empiricism, *rationalism* toward the etiology of delusion is the view whereby delusions are a matter of *top-down* disturbance in some fundamental beliefs of the subject, which may consequently affect experiences and actions (Campbell 2001, p. 89). Rationalism reverses the direction of causal explanation in relation to empiricism, attributing the cause of the delusion to straightforwardly organic factors and, hence, denying that there is any rationalizing explanation of delusions. Top-down approaches have been proposed to account for Cotard (Gerrans 1999), Capgras (Campbell 2001), and for delusions of alien control, or passivity experiences, in which the subject experiences her movements, thoughts or feelings as somehow controlled or generated by an external force (Stephens and Graham 2000).

Campbell has offered a rationalist model motivated by the observation that delusions appear to function as beliefs that structure subsets of beliefs in fundamental ways, in a marked parallel to what Ludwig Wittgenstein referred to as 'framework propositions.' In *On Certainty*, Wittgenstein discusses the epistemological status of propositions like 'There are a lot of objects in the world,' 'The world has existed for a long time,' 'There are some chairs and tables in this room,' 'This is one hand and this is another,' and so on. Such propositions cannot be doubted because they have been made 'exempt from doubt' (§341), functioning like the fixed hinges on which all other considerations turn. They are the end-point of justification: if someone were to express doubt in such propositions, or to ask us to justify our assent to them,

we would immediately question whether they understood the meaning of the words being employed. What is more: we would question their *sanity* (1969, §71, §572).

This invites the comparison of the kind of status that we ordinarily assign to framework propositions with that assigned by delusional subjects to such propositions as 'I am dead,' 'my spouse has been replaced by an impostor,' and so on. Their utter incorrigibility suggests that they should be treated as background assumptions rather than ordinary propositions open to falsification—as constraining the subject's reasoning and interpretation of their experience like a fundamental framework. This kind of analysis of the Cotard and Capgras delusions would lead to an expectation of top-down consequences in the affective aspects of the patient's perceptions of other people, since if you believe you are dead, you will not interact with other people as you would if you were alive, and the same will hold if you think your spouse or anyone else has been replaced by an impostor.

Furthermore, with regard to explanatory advantages over empiricist analyses of Cotard and Capgras, the rationalist analysis predicts that the subject who moves from one set of framework principles to another *destabilizes* the meanings of the terms used. In this way, rationalism provides an answer to an important question left by empiricism, namely, whether the delusional subject can be said to be holding on to the ordinary meanings of the terms used to express the content of the delusion. Campbell likens the shift in the meaning present in such delusions as Capgras and Cotard to the shift in the meaning of terms used in scientific theory before and after a revolutionary change in the key principles of the discipline, as described by Thomas Kuhn (1962). So the meaning that the memory demonstrative 'that [remembered] woman' bears before the onset of Capgras delusion, for example, is unrelated to the meaning that the subject assigned to it after the onset of the delusion—except as a historical antecedent, like 'mass' in classical mechanics in relation to 'mass' in relativistic physics (2001, p. 98).

A consequence of Campbell's rationalism is that incomprehensibility will vary according to the number of shifts in meaning a delusional subject goes through. In monothematic presentations of Capgras or Cotard, for example, only a circumscribed loss of understanding will afflict the interpreter. In polythematic delusional systems such as affect some patients of schizophrenia, in turn, a wide range of the subject's assertions may be consigned to solipsistic meaninglessness (Sass 1994).

Tim Bayne and Elisabeth Pacherie (2004, p. 7) note, however, that it is difficult to reconcile Campbell's thesis that delusional subjects have lost their grip on the meaning of the terms they use in the context of explaining a delusion with the fact that a number of patients, such as the one interviewed

by Michael Alexander and colleagues (1979, p. 335) and quoted above, are able to grasp the fact that others find it difficult to believe their story.

With regard to the explanatory power of Campbell's account, as well as top-down accounts in general, Bayne and Pacherie (2004, p. 8) note that it is puzzling how a delusional framework belief could, in a top-down fashion, cause the damage to the autonomic system seen in the Capgras and Cotard delusions (Ellis et al. 1997; Young 2000)—an important part of the evidence for the bottom-up account of at least these two delusions. To the list of empirical findings that any top-down theorist will be required to take into consideration, Jakob Hohwy (2004, p. 65) adds that it is equally mysterious how belief could, in a top-down fashion, explain decreased ability for fast error correction in schizophrenia (Frith and Done 1989), as well as modulation of activity in parietal cortex in delusions of alien control and other passivity experiences (Blakemore et al. 2003). Also, in schizophrenia, it is not clear how the delusional belief could explain the increased sensitivity to self-produced stimuli, such as tickling oneself (Blakemore et al. 2000) and the sound of one's own voice (Ford et al. 2001).

Whereas neither empiricism nor rationalism on their own seem able to address the necessary conceptual and explanatory needs raised by delusions while heeding the relevant empirical findings, the sheer variety of delusion suggests the unlikelihood that *all* delusions will be accounted for as either exclusively bottom-up explanations of anomalous experiences or exclusively a matter of top-down disturbance. Accordingly, some recent models have proposed that this divide is not unbridgeable and, hence, should be thought of as more of a didactic simplification. In a model of the delusion of alien control in schizophrenia, Hohwy illustrates how top-down and bottom-up processes may coexist in a model of delusion formation (Hohwy and Rosenberg 2005).

> Top-down: hypofrontality[27] associated with posterior hyperactivity modulates the experience of self-initiated movement so that it is experienced the way one experiences externally generated movement.[28] Bottom-up: there is no inhibition of the pre-potent doxastic response (i.e., our tendency to believe what we experience),[29] and the content of the experience is adopted as belief.

[27] A state of decreased cerebral blood flow in the prefrontal cortex during tests of executive function that is commonly observed in schizophrenia (Spence et al. 1998). [28] Movements attributed to an external source result in cerebellar-parietal hyperactivity compared to identical movements correctly attributed to the self (Blakemore et al. 2003). [29] As we have seen, a deficit in belief revision predicted and incorporated in empiricist two-factor frameworks (Davies et al. 2001).

> Top-down performance failure:[30] the beliefs based on experience in some sensory modalities or at some processing stages are inaccessible to reality testing, they cannot then be revised, and are subsequently explained in terms of supernatural hypotheses. The supernatural theme is prioritized because what needs to be explained is the belief that the movement is externally generated, not the belief that it is as if the movement is externally generated, in which case themes concerning the patient's possible mental illness would be more likely to present themselves. (2004, p. 67)

If something like this story is correct for even *one* type of delusion, the generalizations of both empiricism and rationalism are defeated. It strongly suggests that we should ignore the craving for generality and work toward a more nuanced picture of delusion formation that does not prioritize only one direction of causation (at least not for *all* exemplars of delusion).

Finally, every theory of delusion formation discussed so far assumes that explaining delusions is a special case of explaining *beliefs*, however pathological.[31] But characterizing delusions as aberrant beliefs may be seen as overlooking the total psychopathology of delusional subjects and neglecting to take into consideration the experiential character of their delusions. Furthermore, while delusions appear to be belief-like in some ways, they also depart from stereotypical beliefs in other important ways, failing to have the expected connections to reasoning, action, and affect that normal beliefs possess. Hence, philosophical considerations about belief attribution (and self-attribution) will be relevant in determining the doxastic status of delusion.

1.3 The characterization of delusion

1.3.1 Delusional experience in schizophrenia

Consider the following testimony of a patient of schizophrenia:

> I've never rigidly held my beliefs about Pepperidge farms [a brand of baked foods, especially cakes] and microwaves, but they've always involved a strong feeling of fear and aversion, related to my feeling that nothing exists—however, I have acted consistently,

[30] Conceiving of delusions as failures of performance, rather than competence, makes sense of cases where delusional subjects recognize the unusual nature of their belief as well as of cases in which delusions resolve and normal inferential performance returns, suggesting that the neuroanatomical basis of inferential competence is preserved (Gerrans 2001). [31] The question of whether delusion is a kind of belief is taken up in chapter 3 of this dissertation.

> over long periods of time, as if these beliefs were unquestionably true—for many years—until three years ago—I didn't eat any p.f. food (unless I'm very hungry or have a low blood sugar—I'm diabetic)—but I've always had a dimension of doubt about these beliefs, and of course I realize how profoundly irrational they sound to other people—this pattern, configuration of my mind, has led some mental health professionals to regard these (half-)beliefs of mine as obsessions, not delusions, and I see their point—I remember once telling my brother-in-law that I had delusions about earthquakes—he said to me, 'well, if you know that they're delusions, how can they be delusions?'—that played into another half-belief of mine, that I don't really feel these fears, but am instead faking mental illness, and then I castigate myself for my 'evil' nature—then, in another twist of the screw, I remember how seriously and painfully I have taken my delusions/obsessions, and I absolve myself of the 'faking' charge—I would much prefer to believe that I am delusional rather than that all these magical events and processes are real. (Sass 2004, p. 79)

Complex cases such as this seem to beggar description in terms of belief—not only from the outside, but from the inside as well. On the one hand, the patient quoted above expresses a failure to experience his own delusions as full-fledged beliefs. On the other hand, these delusions consistently guide his behavior. Still, as the Capgras patient quoted by Alexander and colleagues (1979), the patient recognizes the irrational character of his delusions. The patient's ambivalence toward the content of his delusions is so great as to elicit doubts in himself as to whether he is somehow 'faking' his own mental illness. And this ambivalence is starkly manifested in the fact that he falls short of unambiguously self-attributing belief in the delusional contents described. Careful consideration of the phenomenological character of such schizophrenic delusions has led Louis Sass (1994) to suggest that delusions are not treated by the patient as representations of how things are but rather as *expressions* of the way the subject experiences the world—even if sometimes these expressions use the language of belief to express the bizarre and disorienting nature of the patient's experiences. In other words, the way the delusional contents are treated by the subject is not the way we treat empirical beliefs. This is perhaps best represented by the feature of schizophrenic patients that Sass calls 'double bookkeeping.'

Consider the following cases: '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' (Sass 1994, p. 21). How is this possible? Sass maintains that the relative internal coherence of the patient's thoughts is safeguarded by his or her keeping two sets of mental "books." In the first book, the one used for everyday life and social interaction and the one which nondelusional subjects share, the patient's thoughts are treated as empirical beliefs subject to reality testing by the use of *intersubjective* standards of confirmation. Moreover, as empirical beliefs, these thoughts will have the appropriate, stereotypical connections to reasoning, action, and affect. Of course, this represents the vast majority of even the most floridly delusional patient's beliefs.[32]

In the second book, in turn, intersubjective standards of confirmation are suspended, as are the usual connections to the patient's other mental states, action, and emotion. In this book, thoughts are treated in an extremely *subjective* fashion (so much so that Sass likens this cognitively unstable attitude to an expression of solipsism). As Jennifer Radden (2011, p. 9) notes, this view was anticipated by Immanuel Kant in his *Anthropology from a Pragmatic Point of View*, who described delusional states as 'a play of thoughts in which he sees, acts, and judges, not in a common world, but rather in his own world (as in dreaming)' (2006, p. 114). With regard to how double bookkeeping might work, Gerrans (2013, p. 86) provides a helpful illustration. A violent headache might trigger the thought 'I have a brain tumor'. In the of case someone who enters this thought in the first (intersubjective) book, that thought is quickly cancelled, since one will consider alternative causes (e.g. 'I banged my head in the kitchen counter earlier today'). However, in the case of someone who enters this thought in the second (subjective) book, the absence of a commitment towards revising or replacing the thought if another has better epistemic credentials will result in its adoption. As Gerrans observes, Sass's conception of delusion 'represents a psychology trying to maintain an unstable solipsistic attitude, which is why the patient has to keep two sets of books but constantly struggles to reconcile them' (2013, p. 86).

Whatever the scientific merit of Sass's idea of double-entry bookkeeping—whether it will be shown to map smoothly onto a theory of cognitive processes or prove to be just a useful heuristic—the fact is that the experiential accounts that inspired it undermine the ascription of beliefs. Does the patient *really* believe that the doctors and nurses are trying to poison her while at the same time eating the food they give her? Serious consideration of the experiential character of schizophrenic delusions has led many to the conclusion that

[32] As an illustration, Gerrans (2013, p. 86) remarks that even the patient who thought he was inhabited by a lizard (Browning and Jones 1988) shared his beliefs about reptiles, scales, claws and cold-bloodedness with the rest of us.

the experiential character of delusions is more important than the doxastic one and, hence, that it is more important for theories of delusion formation to account for this experiential character than to reduce all delusions to pathologies of belief (Gold and Hohwy 2000, Parnas and Sass 2001, Gallagher 2009). Those that take the functional or causal role of schizophrenic delusions to be enough to reject the reduction of delusion to belief states adhere to an argument that takes the following general structure:

1. Beliefs must play a belief-like functional role.
2. Schizophrenic delusions fail to play belief-like functional roles.
3. Therefore, many delusions are not beliefs.

Indeed, this is the most popular argument against the doxastic conception of delusions. The argument's first premise comes from functionalism about belief, the idea that what it takes for a state to be a belief is for it to play a certain functional role, and is rarely questioned.[33] The argument's second premise comes, as we have seen, from attention to first-person accounts and clinical observations. But note that a divide-and-conquer suggestion could be made to the effect that we should recognize the attitudinal and experiential differences between conditions that involve complex delusional systems, such as those held by some patients suffering from schizophrenia, and those held by patients in the grip of monothematic delusions. In this way, we could say that, while polythematic delusions beggar characterization in precise doxastic terms, sufferers of circumscribed delusions such as Capgras, Cotard, and erotomania *do* believe the strange propositions which their verbal and nonverbal behavior often suggests they do.

1.3.2 The functional role of monothematic delusions

Does the functional role of monothematic delusions[34] match that expected of belief? What is the functional role of belief? Its folk-psychological attribution points mainly to its use in reasoning, to its behavior-guidance, and to its effects on emotion. Delusions, in turn, often present a high degree of circumscription (Young 1999, Egan 2009), lacking the holistic character expected of beliefs and failing to respect the notion of a coherent belief system whose

[33] But see Miyazono and Bortolotti (forthcoming) for an exploration of the option of resisting the argument by proposing an alternative theory of belief according to which what is necessary for a state to be a belief is to have the function of playing a belief-like causal role, not to actually play the role. [34] For the remainder of this section, 'delusion' will refer to monothematic delusion.

adjustments to one belief imply adjustments to many others. The first kind of circumscription we may point out is *inferential circumscription*. Delusional subjects often fail to draw the obvious logical consequences of their delusions and show little interest in resolving apparent contradictions between their delusions and the rest of their beliefs. For example, the worldview of most patients with Capgras delusion does not seem to change at all as a consequence of supposedly adopting the belief that their spouses have been abducted and that the person they see in front of them is an impostor (Davies and Coltheart 2000). Whatever this state is, therefore, it is severely encapsulated, failing to be integrated with the subject's web of belief. If we embrace even a minimal consistency constraint on belief-ascription, then, in cases such as Capgras, we will be reluctant to say that subjects genuinely believe the content of their delusions. Indeed, authors such as Willard van Orman Quine and Joseph Ullian (1970), as well as Jerry Fodor (1983), have argued that one of the attributes of a belief *qua* belief is its property of being inseparably connected with other beliefs of potentially widely diverse contents. While one may ascribe false belief to subjects for any number of reasons, a state that fails to have the appropriate connections to the subject's other mental states may thus be seen as falling short of being properly described as a belief.[35]

Likewise, delusions often manifest *behavioral circumscription*. While belief has profound connections to action, some delusional subjects fail to act in ways expected of agents who really believe the content of their delusions. As Gregory Currie (2000, p. 175) observes, delusion exerts a powerful psychological force, absorbing inner mental resources, but it often fails to engage behavior in the way expected of genuine belief,[36] a likely consequence of the inferential circumscription noted above. Young (2000, pp. 53) notes that many people who experience Capgras accept the substitutes with a kind of compliant equanimity, and some are even actively friendly. This is evidence that some delusions also manifest *affective circumscription*. Delusional patients often fail to exhibit the emotional responses one would expect of a person who believes the content of her assertions. This is, of course, coherent with both inferential and behavioral circumscription. The failure of

[35] This is precisely the vein in which Gregory Currie and Ian Ravenscroft, opponents of the doxastic conception of delusion, affirm that 'If someone says that he has discovered a kind of belief that is peculiar in that there is no obligation to resolve or even to be concerned about inconsistencies between these beliefs and beliefs of any other kind, then the correct response to him is to say that he is talking about something other than belief' (2002, p. 176). [36] With respect to schizophrenia, this characteristic inertia was already noted by Bleuler, who stated that his delusional patients 'rarely follow up the logic to act accordingly, as, for instance, to bark like a dog when they profess to be a dog' (1924, p. 144) and that 'none of our generals has ever attempted to act in accordance with his imaginary rank and station' (1950, p. 129).

integration between the subject's delusional state and the subject's emotions is manifest in Capgras patients who are more often than not unmoved by the fate of their relatives whom, according to the doxastic interpretation of this delusion, they believe to have been gone missing or even abducted. For example, the aforementioned patient of Alexander and colleagues 'described positive feelings toward 'both wives,' showed no anger or distress about his first wife's desertion, and specifically expressed thankfulness that she had located a substitute' (1979, p. 335). Moreover, George Christodoulou (1977) found that four out of eleven patients had a strongly positive relationship with the misidentified person, and Geoffrey Wallis (1986) noted in a review of cases that around 30% were friendly toward the duplicates. Even if we accept the sincerity of affirmations such as 'This woman is not my wife' and 'My wife has been replaced by an impostor,' belief attribution is hampered in cases where patients do not go looking for their missing loved ones or call the police to report their missing status, as well as where they do not show any signs of being emotionally affected by their supposed abduction.

Be that as it may, just as there are examples of the failure of delusions to be integrated with the subject's beliefs, actions, and emotions, there are also cases that *do* display integration and, therefore, support the attribution of belief (Bayne and Pacherie 2005). For example, a review of 260 cases of delusional misidentification by Hans Förstl and colleagues (1991) found that physical violence had been noted in 18% of cases. Andrew Young and Kate Leafhead (1996) note that all their Cotard patients displayed at least some measure of congruent behaviors, such as refusing to move, to eat, or to shower. J.M. O'Dwyer (1990) reports that erotomania patients commonly act on the basis of their delusion. And Simon Wessely and colleagues (1993) note that 77% of a total of 59 delusional patients acted on their delusions in the month prior to admission. Therefore, circumscription objections have only the power to undermine the *generality* of a doxastic characterization of monothematic delusions, which is a far cry from establishing the generality of a nondoxastic characterization.[37] As the data we have seen so far points to the highly heterogeneous nature of the category of delusion, and even of the subcategory of monothematic delusion, this should not come as much of

[37] Proponents of nondoxastic characterizations follow two main strategies. The first is to look for another propositional attitude that can match the functional role of delusion better than belief. This can be done by either pointing to an attitude that is already part of our folk psychology, such as imagination (Currie and Ravenscroft 2002); to a hybrid attitude that can do the required work (Egan 2009); or to different *kinds* of belief (Frankish 2009). The second strategy, in turn, is to propose that delusions are second-order attitudes towards beliefs (Currie 2000, Stephens and Graham 2007). Positive alternatives to doxasticism will be sketched in section 2 of chapter 3 of this dissertation.

a surprise.

However, even if we concede the heterogeneity of the category of delusion and attempt to classify delusions as more or less belief-like, and even if we recognize the near complete absence of clear-cut cases, there still remains the possibility that there may not be enough determinacy in our ordinary conception of *belief* for there to be a fact of the matter as to whether certain delusions are genuine beliefs or not (Hamilton 2007). Not only can the disanalogies between delusions and beliefs be seen as justifying the withholding of a clear belief attribution, but the ambition to always arrive at a yes or no answer to the question of whether delusional subjects believe the content of their delusions may ultimately be misguided. In other words, the difficulties I have surveyed concerning the attribution of belief to delusional subjects may be partly due to the lack of clear boundaries in the very concept of belief.

1.3.3 The vagueness of 'belief'

The debate concerning the doxastic status of delusion has been mostly developed on the assumption that, ultimately, there is always an answer to the question whether delusional patients believe the content of their delusions and, if they don't, to the question of what their attitude toward such content precisely constitutes. But are we authorized to rely upon these assumptions? One line of reasoning regarding the attribution of belief in borderline cases suggests that such optimism with regard to the descriptive powers of folk-psychological concepts is unwarranted.

H.H. Price, in his famous series of lectures on belief, discussed the not uncommon phenomenon wherein a person may systematically feel himself to be and act as if he were fully committed to a proposition in one set of circumstances, while systematically feeling and acting as if the opposite were true in others. He called this 'half-belief' (1960/1969, pp. 302-14). More recently, Eric Schwitzgebel has alluded to this kind of variability with context and mood, and observed that there are countless cases like this, in which a simple yes or no answer to the question 'Does S believe that p?' doesn't seem to be available. He calls the state wherein a person is not quite accurately describable as believing a proposition, nor quite accurately describable as failing to believe it, a state of *in-between belief* (2001, p. 76).

Schwitzgebel evokes a familiar example in the same vein as Price's case of the half-believing theist. Price suggests the case of someone who on Sundays bears all the subjective and objective marks of someone who believes that there is a God, but who on weekdays bears none of them. Schwitzgebel, on the other hand, suggests the case of someone who, in certain moods and in certain contexts, bears all the subjective and objective marks, and who, in

other moods and contexts, doesn't. (The latter spectrum may include circumstances from those of weakened confidence, as when someone thinks of God as 'a beautiful metaphor', to those where confidence is removed completely from recognition or memory.) Though he may be a regular Sunday churchgoer, he does not feel the urge to defend himself or his religion when, for example, his atheistic friends mock religious belief. In fact, at such moments (especially on weekdays), he may even find himself mildly convinced of the incongruousness of theistic dogma. How can we decide, then, whether he believes that God exists?

> One might say that his beliefs change from occasion to occasion—that as he is grousing about the church social, he does not believe that God exists; as he is rejoicing in the magnificence of spring, he does believe—but most of the time he is doing neither: he is eating breakfast or mowing or writing code and not giving the matter any thought. At such moments he may be simultaneously disposed to marvel at the wonder of creation if a robin were to fly past and to embrace atheism if Madge were unexpectedly to drop by. (Schwitzgebel 2001, p. 78)

In the most difficult cases for ascription, the communicative demands on the attributor may not successfully determine whether or not it is appropriate to describe the subject as believing, say, the content of their delusion. Cases like these, in which the set of ascribable dispositions available to the interpreter is such a "mixed bag," seem to leave us only with the option of *specification*—that is, describing how the subject's dispositions conform to the stereotype for the belief in question and how they deviate from it. There will be times, then, when withholding the use of ascriptive language is going to be preferable so as not to mislead one's audience. Such cases are those in which the observable deviations raise questions regarding both the content of the subject's attitude and the nature of the attitude itself. In the context of the discussion of how best to characterize delusions, Schwitzgebel has recently proposed that, if there is no way to decide whether something is determinately a case of belief, our move should be to allow *some* indeterminacy in our belief talk. Schwitzgebel suggests that 'believes that p' should be treated as a vague predicate admitting of vague cases:

> In in-between cases of canonically vague predicates like 'tall', the appropriateness of ascribing the predicate varies contextually, and often the best approach is to refuse to either simply ascribe or simply deny the predicate but rather to specify more detail (e.g., 'well, he's 5 foot 11 inches'); so too, I would argue, in in-between cases of belief. (2012, p. 15)

Lisa Bortolotti (2010, pp. 20-1) dismisses this, which she terms the 'sliding scale' approach, on the grounds that such an approach, by not giving a straightforward answer to the question 'Does the delusional patient believe that p?', is unable to characterize precisely whether the patient's actions are intentional, which complicates issues of ethical and policy-guiding import, etc. However, apart from this not being nearly enough reason to discard the approach, its proponents might just as well suggest, as Schwitzgebel does, that 'in many cases of delusion it *shouldn't* be straightforward to assess intentionality, and that the ethical and policy applications *are* complicated, so that a philosophical approach that renders these matters straightforward is misleadingly simplistic' (2012, p. 15).

So the conclusion is not quite that, say, the Capgras patient *doesn't* believe that her loved one has been replaced by a double, or that the Cotard patient *doesn't* believe that she is dead. Rather, it is that the question as to whether these subjects believe the content of their delusions cannot be answered plainly—which doesn't mean we should give up our efforts to understand delusion, but that we should shift our attention to what we *can* do, that is, we should attempt to characterize and explain delusion in the levels of description wherein it can be precisely accounted for. As Graham (2010, p. 337) observes, delusions are messy, compound, and complex psychological states or attitudes, defined more by how persons mismanage their content and fail to prudently act in terms of them, than by qualifying as beliefs. Moreover, a realistic picture of delusion should allow for the clinical variation of delusional presentation and not try to funnel each case of delusion through the taxonomic filter of the propositional attitude of belief. The prolonged debate over how to characterize delusional states is predominantly due to its participants using folk-psychological tools that simply may not be up to the task.

Conclusion

In the preceding sections, I have attempted to elucidate at least two facts: first, that 'delusion' is a highly ambiguous term; second, that the phenomena to which it refers are multi-faceted. Moreover, through the examination of the many problems involved in determining the nature of delusions, as well as how they should be explained scientifically and characterized in folk-psychological terms, I have tried to show some of the reasons why philosophers have been progressively disregarding disciplinary boundaries and contributing to the many debates discussed above. Especially, I have made an effort to demonstrate that the engagement of philosophers with the clinical literature on

delusion, and the collaboration between philosophers and psychiatrists, is a two-way street. While philosophers profit from psychiatry inasmuch as the clinical literature provides real-life, as opposed to merely imaginary, cases for philosophy of mind to engage with, philosophers can contribute not only by clarifying concepts and working out the implications of empirical results, but also in building explanatory models of delusion and suggesting new avenues for empirical research. The best way for philosophers to contribute to the understanding of the relevant phenomena, I suggest, is for us to heed Sass's (2004, p. 71) advice and resist the tendency to formulate issues and arguments in overly polarized terms and then to rely uncritically on these formulations in exploring the domain of inquiry, so as not to actually hinder our understanding of phenomena which are often fraught with ambiguities and complexities that defy standard conceptualizations.

Chapter 2

The natural kind status of delusion

Introduction

Delusion is defined by the most recent edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) as a 'false belief based on incorrect inference about external reality that is firmly held despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary' (American Psychiatric Aassociation 2013, p. 819). Predictably, a great variety of phenomena are apt to be grouped under such a definition. Indeed, people who are deemed to be clinically delusional affirm many different things in many different contexts. Here are some of them (Davies and Coltheart 2000, p. 1):

- 'My closest relatives have been replaced by impostors.'
- 'I am dead.'
- 'I am being followed around by people who are known to me but who are unrecognizable because they are in disguise.'
- 'The person in the mirror is not really me.'
- 'A person I knew who died is nevertheless in the hospital ward today.'
- 'This arm [the speaker's left arm] is not mine, it is yours; you have three arms.'
- 'Someone else is able to control my thoughts.'

- 'Someone else's thoughts are being inserted into my mind.'[1]

What follows is an investigation about our warrant for grouping such disparate phenomena together. My primary aim is to assess the prospects for a scientific theory of delusion through the examination of the scientific respectability of this psychiatric category—a status which is arguably put in jeopardy by the fact that the detection and attribution of delusion seem to stem not from causal classification but from the application of what we may call 'folk psychiatry'. I will do so by first introducing the philosophical notion of *natural kind* and examining the question of whether psychiatric kinds as a whole meet the demands required for a kind to be an objective, mind-independent distinction in nature. I will then introduce a liberal sense in which biological taxa as well as psychiatric categories might be viewed as natural kinds—namely, the *homeostatic property cluster* model. Subsequently, I will introduce and assess how models of the detection and attribution of mental disorder may impact even a liberal understanding of delusion as a natural kind. Finally, I will conclude by making a case for a folk-psychological understanding of 'delusion' in general while also recommending a natural-kind methodology for the investigation of subtypes of delusion.

2.1 Kinds of kinds

Are mental disorders real? One of the main theoretical challenges for psychiatry is to determine whether the *kinds* it investigates are *natural*. Psychiatry's scientific credentials came under heavy criticism in the 1960's and 1970's—the most radical embodiment of which was represented by the so-called anti-psychiatry movement, which questioned whether mental disorder represents the pathologizing of normal problems of living. Thomas Szasz, the father of anti-psychiatry, argued not only that mental disorder as a kind fails to pick a real distinction in nature, but that it is just a 'convenient myth' (1961, p. 113). This intuition is reinforced by controversies such as that over the recent removal of the "bereavement exclusion" in the diagnosis of depression in the DSM-5. Likewise the proposed addition of 'persistent complex bereavement disorder' in an attempt to classify those who are significantly impaired by prolonged grief symptoms for at least one month after six months

[1] These examples pertain to eight different subtypes of clinical delusion, respectively: Capgras delusion, Cotard delusion, Frégoli delusion, mirrored-self misidentification, reduplicative paramnesia, somatoparaphrenia, thought control, and thought insertion. See chapter 1 of this dissertation for a more in-depth introduction to delusion.

of bereavement.[2] Against the backdrop of challenges to the validity of psychiatric classifications as a whole, the task is to make clear the basis on which conditions are included or excluded from the manuals and why this basis is scientific and objective and not just a matter of social rules of normal behavior (Bolton 2008, p. 164). If entities classified as mental disorders could be shown to be natural kinds, then many of the controversies surrounding the status of psychiatry as a serious scientific endeavor could be resolved. However, this will depend on what exactly one takes natural kinds to be.

2.1.1 Essentialism about natural kinds

What are natural kinds? What characteristics must a kind have in order for it to be considered a natural kind? The traditional account of natural kinds is represented by various forms of *essentialism* which date back to the Aristotelian tradition, in which essences had both causal and classificatory (sortal) roles. The causal role referred to the underlying properties that determined and sustained an instance's visible properties. Because these underlying properties were supposed to be fixed, they were identified with the nature of a kind—that which makes it be what it is. After the rise of natural philosophy in the seventeenth century, the essential hidden properties which Locke called 'real essences' came to be identified with underlying structural properties which, he argued, are not observable.[3] In the twentieth century, essentialism was mostly related with the revival of the notion of natural kinds in the work of Saul Kripke (1972) and Hilary Putnam (1975),[4] which followed the skepticism about the stability of scientific knowledge brought about by the work of Thomas Kuhn (1962).

As Marc Ereshefsky (2009) observes, essentialism usually involves three main tenets: first, all and only the members of a kind share a common essence; second, that essence is a property, or a set of properties, that all the members of a kind must have; and third, a kind's essence causes the other properties associated with that kind. So, for example, the essence of gold is gold's atomic structure, and that atomic structure occurs in all and only pieces of gold. That structure is a property that all gold must have as opposed to such accidental properties as being valuable to humans. And the atomic structure of gold

[2] Persistent complex bereavement disorder was placed in the chapter 'Conditions for Further Study' in the DSM-5 after its proposed addition generated a great deal of controversy.
[3] It is fair to say that Locke underestimated the kinds of observation that technology would eventually allow us to make of properties which are potentially essential, such as the number of protons in the nucleus of an atom, or the genetic code in specific DNA sequences.
[4] But see Hacking (2007) for criticism of the lumping together of Putnam's and Kripke's theories of natural kinds on the basis that Putnam was not an austere essentialist.

causes pieces of gold to have the properties associated with that kind, such as readily dissolving in mercury at room temperature, conducting heat and electricity, and being unaffected by air and moisture.

The reason why it matters for the development of a science that its kinds be natural in the sense of picking up essential distinctions has to do with the fact that such kinds will be ideally suited to figure in key scientific practices such as induction, explanation, classification, and discovery. Natural kinds pick out classes about which non-accidental, scientifically relevant, inductive generalizations can be formulated, since its members share many non-accidentally related properties. The reliably co-varying clustering of properties that instances of natural kinds possess is, however, contingent (as opposed to logically or conceptually necessary) and its existence calls out for explanation, usually undertaken through the identification and specification of the structures, processes, and mechanisms that causally explain the property clusters associated with the kind under consideration.

In other words, one's ability to make inferences about members of a natural kind is explained with reference to their shared underlying properties. Being some such natural kind explains why an instance of that kind has the features that it does, and that explanation is to be found in studying the intrinsic underlying properties an instance shares with other instances of that kind. Furthermore, with respect to the classificatory role, if one can identify the essence of a thing, one may be able to determine its place in the natural order. According to essentialism, if you want to know whether something is a true member of a natural kind, you should check whether the causally essential underlying properties are present, as such properties will invariably be necessary and sufficient conditions for membership in a natural kind. Thus, essentialism implies that there is a correct classification of naturally occurring kinds out there waiting to be discovered. As the philosophical adage goes, nature is such that it can be "carved at its joints."

Besides figuring in the practices of generalization, explanation, classification, and discovery, Richard Samuels (2009) points out three further characteristics that flow from natural-kindhood as necessary conditions for the scientific respectability of any given kind. Given that natural kinds possess a *sortal* essence,[5] they will be *discrete* classes of entities that can be clearly demarcated from other phenomena and they will be highly *homogeneous* classes as well. Moreover, natural kinds will be *mind-independent* in an important

[5] As Samuels (2009, p. 57) uses the term, sortal essences consist of intrinsic properties and, as a matter of metaphysical necessity, they are possessed by all and only the members of the kind. *Causal* essences, on the other hand, do not imply these commitments, and are simply the set of properties that figure in causal explanations of a given kind. So all sortal essences are causal essences but not vice versa.

sense,[6] which Sam Page (2006) calls *individuative independence*, namely, that of being circumscribed by boundaries that are totally independent of how we categorize things. Page illustrates his concept by alluding to the individuation of the night sky into constellations: 'Though it is *prima facie* plausible that reality is individuated intrinsically into stars, reality is not individuated intrinsically into constellations, since it is people who divide the night sky into constellations' (2006, p. 328).

Essentialism about psychiatric kinds—the view that psychiatric disorders are (or at any rate should be) akin to stars, not to constellations—is associated with the biomedical model of psychiatry, which proposes that psychiatric kinds can and should be isolated by studying underlying biopathological processes. Jerome Wakefield's (1992) *harmful dysfunction* model, arguably the most important philosophical theory about the nature of mental disorder, recognizes the claims of Szasz and others concerning the evaluative nature of psychiatric diagnosis without thereby abandoning realism about psychiatric disorders. Wakefield argues that the presence or absence of a dysfunction is a factual matter, just as the presence or absence of a natural function is. Since natural functions were selected for during evolution because of their contribution to the survival of the organism, evaluative statements about functions (and, hence, dysfunctions) can be translated into objective, factual statements about evolutionary history. To qualify as a "disorder," however, Wakefield acknowledges that there must also be evidence that the condition in question is *harmful* to its bearer—and this will be an inherently evaluative, normatively assessable aspect of all judgments of pathology.

Given the present stage of development of biological psychiatry, however, the essences of the dysfunctions that constitute psychiatric disorders—alongside the evaluative aspect of suffering or impairment—are yet to be discovered, just as the essence of electrons and gold once were. Until the necessary scientific discoveries are made, their essences are, so to speak, in a black box. As Peter Zachar explains, Wakefield's (2004) black-box essential-

[6] Following Page (2006), Samuels (2009, pp. 53–4) identifies three possible senses of mind-independence that do *not* flow from natural-kindhood and are, therefore, irrelevant to the characterization of natural kinds. The first is that attached to theoretical entities (e.g. quarks, electrical fields, and chemical compounds), which should not be considered trivially mind-dependent, non-natural kinds. The second is that attached to entities whose existence metaphysically necessitates the existence of minds, such as psychological kinds as beliefs, desires, delusions, etc. and, again, should not be considered trivially non-natural. Finally, and perhaps more controversially, Samuels rejects the relevance of causal dependence on mental activity, which is true of such kinds as toy poodles and the radioactive chemical element californium, as he argues that this feature should not trivially imply that such kinds are not "natural" in the *scientifically* relevant sense (i.e. though not naturally-occurring, they may nevertheless turn out to figure in all relevant scientific practices).

ism follows the scenario proposed by Putnam and Kripke wherein, at some point in history, there occurs a "baptismal" event in which, in the example at hand, a disorder is clinically observed and named: "This is psychopathy,' said Hervey Cleckley (1941). 'This is autism,' said Leo Kanner (1935). If the original disorder concept can be developed into a proper scientific construct (one based on an objective dysfunction), the clinician's original concept can be said to have indirectly referred to the objective dysfunction all along' (2014b, pp. 83–4).

Note, however, with respect to the aforementioned conditions for the scientific respectability of a kind, that biological taxa such as species appear to meet all of them and, still, they are widely regarded as failing to constitute essentialistic natural kinds[7] as do chemical kinds such as ascorbic acid and H_2O, and physical kinds such as quark and lenticular galaxy. This is the case because, as the first tenet of essentialism requires, for a biological trait to be the essence of a species that trait must occur in *all and only* the members of that species. However, as Ereshefsky (2001, p. 98) points out, a number of biological forces work against the uniqueness and universality of a trait in any given species. For example, suppose a genetically-based trait were found in all the members of a species, such as the unique genetic code of lemons that Putnam (1975) speculates is the essence of lemons. The forces of non-adaptive causes of evolution such as mutation and genetic drift can cause the disappearance of that trait in a future member of the species. Furthermore, as Ereshefsky observes, even if a trait occurred in all the members of a species, that trait would be the *essence* of a species only if it were unique to that species. But organisms of different species often have common traits because they inherit similar genes and developmental resources from common ancestors. Therefore, given the requirements of essentialism and the forces of evolution, essentialism about biological kinds has been widely rejected.[8]

If biological kinds are not amenable to conceptualization as natural kinds, then what chance do psychiatric kinds stand of successfully being characterized as such? Zachar (2000) argues that conceptualizing psychiatric disorders as bounded entities in nature is inconsistent with evolutionary biology's understanding of species. Indeed, as Nick Haslam (2014, p. 11) notes, psychiatric

[7] From now on, I drop 'essentialistic' as always refer to natural kinds in the essentialistic sense unless otherwise noted. As we will see below, the term 'natural kind' has been re-appropriated by authors who believe that essentialism is too stringent, while believing that less stringent criteria can properly characterize kinds as 'natural' (Boyd 1991). [8] Three main views have been advanced in response to this: denying that species are natural kinds and looking elsewhere in biology for kinds with essences (Hull 1978); arguing that species are indeed kinds with essences, but that their essences are of a non-traditional variety (Okasha 2002); and, as we will see below, arguing that natural kinds do not require the sort of essences implied by essentialism (Boyd 1999).

classification would be a great deal easier if its diagnostic entities were like biological species, since, while the process of demarcating biological taxa rests on the scientifically impeccable confidence that naturally occurring biological kinds exist, the taxonomic situation in psychiatry is very different, as mental disorders do not pick out distinct, reproductively isolated, spatially concentrated populations. Moreover, while biological species are "indifferent kinds", at least some mental disorders seem to be "interactive kinds" (Hacking 1999), since those who are classified are often aware of being labeled and may come to change their behavior and even their self-experience in consequence of such awareness, thus producing a "looping effect" whereby the labels may change in virtue of their subjects changing (Hacking 2007b).

Furthermore, in stark contrast to their biological counterparts, psychiatric kinds (and *kinds of people* more generally) tend to be at least partly shaped by social processes and normative concerns. These considerations are the motivating force behind the anti-essentialist argument in philosophy of psychiatry. As we will see, the cogency of this argument will depend on how exactly one should understand 'essence', as essentialism about natural kinds has been challenged in recent years (Boyd 1991). Also, it will depend on the plausibility of the repudiation of pluralism—the view that different psychiatric kinds differ in how much they fail to meet the criteria for natural-kindhood (Haslam 2002)—the acceptance of which would in principle keep open the possibility that at least *some* mental disorders might have essences. For now, however, I will assume that the general argument is cogent in order to consider what may be proposed instead to properly capture the features of psychiatric kinds, noting that by assuming that they are not natural kinds one is not immediately committed to the view that they are non-kinds (*pace* Szasz).

Following a nuanced classification of kinds of kinds, such as that offered by Haslam (2014), will go a long way toward disabusing one of the notion that distinctions proper must be essential or fail to be real distinctions at all. His schematic account is based on five kinds of kinds that satisfy increasingly stringent criteria, each successive kind of kind having to meet one more requirement, with natural kinds being on the top of the ladder.

	Criterion				
Kind type	Clustered properties	Non-arbitrary cutpoint	Discontinuity	Category boundary	Category essence
Dimension	✓	✗	✗	✗	✗
Practical	✓	✓	✗	✗	✗
Fuzzy	✓	✓	✓	✗	✗
Discrete	✓	✓	✓	✓	✗
Natural	✓	✓	✓	✓	✓

Table 2.1: Schematic account of the kinds-of-kinds model (Haslam 2014)

In the remainder of this section, I will go over the different kinds of kinds that fall short of being distinguishable by a category essence: dimensions, practical kinds, fuzzy kinds, and discrete kinds. I will connect these notions to the discussion of natural-kindhood in the philosophy of psychiatry, as well as to the more general discussion of the proper way to characterize natural kinds, within which the most widely adopted view states that natural kinds should not be conceptualized essentialistically, but in terms of property clusters sustained by complex, mutually reinforcing networks of causal mechanisms.

2.1.2 Dimensions and practical kinds

The first kind of kind and the least demanding structure in Haslam's model is what he refers to as *dimensions* (strictly speaking a non-kind, since they do not define delimited categories). The label comes from the standard categorical/dimensional distinction in psychopathology research and theory, motivated by the categories of personality disorder which, perhaps more than any other current DSM category, do not seem to be distinct species (Clark, Watson, and Reynolds 1995; Livesley 2003; Widiger and Sanderson 1995). Zachar (2014, p. 93) alludes to a model introduced by Livesley (2003), in which once the pathological dimensions have been identified—which may include narcissism, impulsivity, anxiousness, social detachment, and hostility (Widiger, Livesley, and Clark 2009)—patients meeting criteria for a broad category called 'personality disorder' are distinguished from one another by their respective position on the dimensions. To qualify as a dimension, all that is required for a kind, such as any given mental disorder, is that there be a set of correlated properties, such as symptoms. As Haslam puts it, 'Individuals may differ by degree along a dimension by possessing greater or lesser numbers or degrees of these properties. Variation along a dimension is continuous and seamless, so there is no naturally occurring break separating individu-

als who are affected with a condition from those who are not' (2014, p. 14). In other words, if psychiatric kinds were dimensions, this would amount to there not being delimited conditions at all. A cutpoint would be defined on the dimension so that the quantitative variation would be simplified into a dichotomous diagnosis, but its placement would be arbitrary.

Thus, proponents of dimensional models of psychopathology hold that the distribution of variation on psychopathology-related dimensions is continuous in the same sense as what philosophers refer to as 'vague predicates'. These models are devised in response to the limitations of the purely categorical approach, such as the failure to capture individual differences in disorder severity, and clinically significant features subsumed by other disorders or falling below conventional DSM thresholds (Brown and Barlow 2005). Nevertheless, while rejecting the view that psychiatric kinds are natural kinds, Zachar (2000) argues that mental disorders pick out reasonably stable, non-arbitrary patterns that can be identified with varying levels of reliability and validity, and that the application of many of the distinctions of psychopathology is justified by its usefulness for clinical purposes, being demarcated on the basis of external considerations rather than on the basis of internal discontinuities. In keeping with these observations, Zachar proposes that mental disorders be conceptualized as *practical kinds*, the next rung in Haslam's ladder, which refers to the least demanding sort of non-arbitrary cutpoint—that of pragmatically grounded distinctions.[9]

As an example from outside the field of psychiatry, Zachar (2014b, pp. 154–5) alludes to the distinction between an adult and a child. Although the kinds 'adult' and 'child' are not in themselves sharply demarcated, the uses for which we deploy them will determine where their boundaries should be drawn. Consequently, many distinctions between adults and children are context-dependent. For example, if our aim is to decide who is able to vote, engage in consensual sex, get married, be sent to prison, drink alcohol, or enter into a legal contract, each of those considerations will result in different ways of demarcating adulthood (Horwitz and Wakefield 2012, p. 53). As medical examples of non-arbitrary cutpoints on continuous dimensions, Haslam (2014, p. 14) points out blood pressure values for diagnosing hypertension and Body Mass Index values for diagnosing obesity—values that roughly correspond to levels at which health risks become more likely. When at some point along a dimension the severity of the relevant symptoms becomes clinically significant or a source of functional impairment, the existence

[9] Though, as we will see below, Zachar's most recent proposal acknowledges the middle way between practical kinds and essentialism about natural kinds embodied in Richard Boyd's property-cluster approach, going so far as to state that Boyd's model is probably the most appropriate for conceptualizing most psychiatric disorders (Zachar 2014, p. 94).

of a non-arbitrary, pragmatic distinction is justified.

So practical kinds, while fuzzier than natural kinds, are not open to the charge of arbitrariness as dimensions are (at least as conceptualized in Haslam's model). The classification of practical kinds requires balancing criteria that do change their values in different contexts depending on treatment goals, research priorities, and disciplinary standards of validity. As a consequence, practical kinds fall short of possessing the perfect reliability one may be justified to expect from natural kinds. Relating the practical-kinds model to his claim that psychiatric nosology is inherently goal-oriented, Zachar has recently elaborated on the dynamics of classification within his model, observing that it emphasizes that discovery of fact contributes greatly to progress in classification, but that discovery alone cannot tell us how to classify: 'For example, discovering that a mild form of cognitive disorganization (schizotypy) is common in families of people with schizophrenia was an important finding that highlighted an objective feature of the world. Should schizotypy, therefore, be classified as mild manifestation of a unitary schizophrenic spectrum (a genetic grouping)? Another possibility is that should it be classified as a premorbid personality style that represents a vulnerability to the mental illness of schizophrenia. In which box should it be placed?' (2014, p. 90). Zachar's point is that, apart from goals relating to classification and theory-building, neither demarcation is privileged in and of itself.

The presence of goal-oriented cutpoints raises the question of whether practical kinds are apt to count as scientifically relevant kinds, and this, in turn, raises the question of the minimal criteria of scientifically-relevant kindhood. Zachar defers to Nelson Goodman, who did not advocate for natural kinds or scientific realism, but instead offered a theory of relevant kinds. With respect to the criteria for relevance, according to Goodman, good scientific kinds *support induction* (to a greater or lesser degree) or, as he would later put it, they have properties that are "projectible," meaning that if we observe certain properties in a subset of a kind, we can infer that these properties will occur in other instances of the same kind, allowing us to confirm generalizations about that kind (Goodman 1978, 1983). Let us assume, for the sake of the argument, that projectibility is a good enough criterion of relevance. Do psychiatric kinds support induction? Even though present classifications of mental disorders are highly variable with respect to validity, and in spite of

diagnosis being presently based on polythetic categories,[10] research on mental disorder has been able to produce many useful generalizations.[11] The question is whether these generalizations are based on (at least some) psychiatric kinds being held together by shared *causal* mechanisms or if they are based solely on these kinds's shared surface features, meaning that they are merely practical kinds.

The practical-kinds model is implicit in the symptom-based nosologies of current diagnostic manuals which aim at grouping patients into useful classes that serve practical goals (such as predicting behavior, assessing genetic risk, or selecting a course of treatment). This grouping, effective as it may be, does not require that diagnoses be grounded in shared causal processes. On the other hand, the assumed causal heterogeneity of psychiatric kinds does not immediately imply that they cannot be causally classified. Note, however, that as the existence of shared causal mechanisms underlying mental disorders is currently an open question, assuming that a causal classification of psychiatric kinds is tenable is something of a "black box" approach (as is Wakefield's harmful dysfunction model). Nevertheless, as Kenneth Kendler, Peter Zachar, and Carl Craver (2011) argue, by focusing solely on the adjustments and compromises that actually occur in classification, the practical-kinds model fails to suggest a way toward progress. In other words, the model is purely descriptive of the current state of psychiatric classifica-

[10] Polythetic (as opposed to monothetic) categories were introduced in the DSM-III (APA 1987) and are still used in the present edition, DSM-5 (American Psychiatric Association 2013). Polythetic classification is carried out by assigning a certain number of criteria, of which some, but not all, need to be met in order for an individual to be a member. So, for example, the diagnosis of schizophrenia is partially dependent on the patient showing two or more of the following symptoms (for much of the time during a one-month period): delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms such as blunted affect, alogia, and avolition (American Psychiatric Association 2013). [11] For example, with respect to depression, preventive efforts result in a decrease in rates of the condition of between 22 and 38% (Cuijpers et al. 2008), and stepped-care intervention (watchful waiting, cognitive behavioral therapy, and medication in some cases) has achieved a 50% lower incidence rate in a patient group aged 75 or older (van't Veer-Tazelaar et al. 2009). With respect to schizophrenia, a combination of new medications and community-case management—a multidisciplinary team of mental health professionals who engage with the patient and their carers inside and outside the hospital, and ensure a combination of health and social care—has resulted in remission of about 80% of patients, especially if treatment is initiated early during the first episode of the illness (van Os and Kapur 2009). With respect to bipolar disorders, prodromal symptoms (i.e. those preceding a relapse) can be reliably identified by at least 80% of individuals with bipolar disorder (Jackson, Cavanagh, and Scott 2003), and teaching patients coping strategies to employ when noticing the symptoms, such as stimulation reduction and seeking professional help, has been correlated significantly with better social functioning (Lam and Wong 2005).

tions. If progress is to be made, however, linking disorders to their etiology and underlying mechanisms is indubitably psychiatry's best bet. For this reason, psychiatry may profit from conceptualizing its kinds in a way that goes beyond the merely pragmatic and assumes internal (but not necessarily external) discontinuities. To this end, we may climb one more rung in Haslam's ladder, toward a more ambitious model.

2.1.3 Fuzzy kinds and discrete kinds

Dimensions and practical kinds both represent forms of continuous variation. According to Haslam, such variation becomes categorical in a deeper sense when there exists some sort of internal discontinuity within a kind which cannot be accounted for by pragmatic considerations alone: 'Such a discontinuity involves a break on the underlying continuum, which produces a qualitative distinction between people who fall above the discontinuity and those who fall below it. An example is a threshold effect, in which a qualitative change of state occurs at a certain point on an underlying continuum (e.g., a liquid turning to a gas at a certain temperature, or a spring losing its tension beyond its elastic limit)' (2014, p. 15). When internal discontinuities within a kind are present but are not sharp, we have what Haslam calls *fuzzy kinds*. Within these, then, kind membership will not always be definite: there will be a penumbra of intermediate cases between those that are definitely members of the kind and those that are definitely not.

On the other hand, when internal discontinuities are sharp but no set of essential properties exists, we step up Haslam's ladder once again to find what he calls *discrete kinds*. In this kind of kind we have what may properly be called a category boundary. However, Haslam points out that discrete kinds may have a variety of possible causal underpinnings, as many types of causal explanation can yield category boundaries: 'These causal explanation types include sharp threshold effects (where the qualitative change of state is abrupt), dynamic interactions of multiple causal factors, and explanations that invoke centripetal tendencies within categories (e.g., conscious identification with a group or label) and/or differentiating tendencies' (2014, p. 15). This immediately makes discrete kinds excellent candidates for scientific respectability in the eyes of those who argue that scientific practice does not require an essence in the traditional sense of a microstructural property that explains all the other properties of a kind while also being unique to that kind.

Indeed, both fuzzy and discrete kinds are candidates for natural kindhood if one refuses to accept that what makes a kind a natural kind is its possession of an essence, rather than its utility in induction and other scientific practices.

Within the non-essentialist *kinds-in-science* tradition (Cooper 2013), fuzzy, discrete, and essentialistic natural kinds are all proper subsets of inductively useful kinds.[12] Within this tradition, several accounts of kinds have been developed with the aim of explaining how it is that kinds like biological species—in which there simply are no essential properties to be found—can successfully ground explanations and inductive inferences. Insofar as the most ambitious sense in which psychiatric kinds might turn out to be natural is the same in which biological kinds are taken to be natural, such accounts of kindhood are of particular interest for the conceptualization of mental disorders as something belonging between practical kinds and kinds with essences.

John Dupré (1981, 1993) argues for *promiscuous realism*—the view that there are countless, yet legitimate ways of dividing up the world into kinds. He asks us to consider the entities of some domain mapped into a multidimensional space wherein the different dimensions map onto different properties, as in cluster analysis—a statistical method for grouping sets of objects based on their similarities, in such a way that objects in the same cluster are more similar to each other than to those in other clusters. According to Dupré, biological species—as well as higher taxa such as families and kingdoms, and lower ranks such as subspecies and varieties—would be identified with some such clusters. His realism has to do with the fact that he accepts that the world possesses individuals which are objectively similar to each other, sharing properties and, thus, being identifiable as being of the same kind. The promiscuity of Dupré's realism, on the other hand, has to do with the fact that he denies that these properties are intrinsic properties of kinds and, in line with Haslam's concept of fuzzy kinds, he argues that natural kinds are not necessarily categorically distinct (i.e., they are not necessarily discrete kinds). Moreover, such taxonomic promiscuity is reflected on our classificatory practices both in the context of common sense and within science.

In the context of common sense, a (presumed) natural kind such as lilies is classified as a flower, although, in biology, species which are commonly referred to as lilies occur in numerous genera of the lily family (Liliaceae), including bulbs such as garlic and onions. However, as Dupré observes, to include the onions and garlics in the reference of the English word 'lily' would surely amount to a debasement of the term (1981, p. 74). The moral is that common sense and biology provide us with pluralistic ways of classifying

[12] Though the history of natural kind thought is usually traced back to Locke's real essences (Boyd 1991), Murphy (2006, p. 335, fn. 6) notes that, as a historical precedent for the kinds-in-science tradition, Hacking (1991) argues that the notion of natural kinds indubitably surfaces in Mill and Venn in the mid-nineteenth century in connection with induction—something which did not preoccupy philosophers before Hume.

lilies and each is equally legitimate depending on our interests. This is not to say that Dupré's kinds are merely practical—it means that his conception of natural kinds takes seriously the different classifications that arise from a variety of interests. Indeed, cross-classification sometimes occurs within the context of a single science, to which the countless ways of classifying species bear witness (Dupré 1993, p. 38).

By denying that there is one unique way of demarcating the set of natural kinds, Richard Boyd (1991, 1999) endorses promiscuous realism. Furthermore, by emphasizing that members of a kind share properties for a reason, his *homeostatic property cluster* (HPC) account elaborates on Dupré's idea. In a near-consensus in recent philosophy of science, the HPC account has been widely seen not only as the most successful approach to make sense of the intuitive natural-kindhood of biological species, but as quite simply the best account of natural-kindhood (Samuels and Ferreira 2010). The HPC model defers to the kinds-in-science tradition by stating that natural kinds are scientifically relevant kinds and that these are, at a minimum, fuzzy sets defined by homeostatic[13] mechanisms at multiple levels that act and interact to produce the key properties associated with the kind. These mechanisms are the reason why members of a kind are, and continue to be, alike. Importantly, they are also the reason why the clusters of phenomena identifiable as being of the same kind are similar enough to be subject to explanation in terms of the same underlying causal properties. Thus, Dominic Murphy (2006, p. 338) refers to Boyd's account as a refined form of essentialism, since homeostatic properties substitute and play the same role of what in "simple" essentialism constituted the essence of a kind (namely, microstructural properties). By not insisting on necessary properties or a single, essential cause, and by not specifying that such a cause must be biological, the HPC account is clearly broader than simple essentialism and advances a much more liberal sense of natural-kindhood.

So Boyd's natural kinds are, minimally, fuzzy kinds. In cluster-analytic terms, if the members of different fuzzy kinds whose members share a certain number of properties are plotted in a multidimensional space, there will not always be a clear gap between them. As Haslam (2014, p. 18) notes, since homeostatic mechanisms merely produce correlations among properties and resemblance among entities that possess those properties, there is no reason to assume that similarity-generating mechanisms will always yield sharp discontinuities between entities that possess sufficient levels or numbers of those properties and entities that do not. This leads Carl Craver

[13] Homeostasis being the property of a system or mechanism by which variables are regulated so that internal conditions remain stable and relatively constant.

(2009) to conclude that HPC kinds have a *prototype* or *family-resemblance* structure. Note, however, that both discrete and essentialistic kinds are also proper subsets of the set of HPC kinds, so that Boyd's account accommodates the intuitively plausible possibility that there are different levels of natural-kindhood—in Haslam's five-tier classification, these levels comprise all kinds for which there are internal discontinuities independent of our interests. In this way, some scientifically relevant kinds may turn out to be fuzzy, others discrete, and still others may turn out to have essences. For example, membership in the kinds encompassed by chemical elements may be essentially defined by the number of protons found in the nucleus of an atom. This is part of the appeal of Boyd's account, since there is no reason to think that psychiatric disorders, biological species, and chemical elements must pertain to the same kind of kind, and, according to the kinds-in-science tradition, there is also no reason to deny natural-kind status to non-essentialistic kinds as a matter of principle.

The inductive potential of HPC kinds is underwritten by the fact that if properties are held together homeostatically, then we will be able to conclude on the basis of one property that others will typically occur with it. Boyd's focus on the underlying causal mechanisms that make homeostasis possible is important for the present investigation because it ties the HPC model to causal explanation and classification which, as we have seen, is absent from the practical-kinds model—the main competing model of psychiatric kinds. As Samuels notes, for any homeostatic property cluster 'there is some set of empirically discoverable causal mechanisms, processes, structures, and constraints—a *causal essence*, if you will—that causally explains the co-variation of these various symptoms' (2009, p. 55). Therefore, kind-membership will be defined not by sets of co-occurring properties or symptoms, as mental disorders are presently demarcated in diagnostic manuals such as DSM-5, but by the set of causal mechanisms that make these properties occur together. On the other hand, psychiatric conditions could satisfy the requirements of an HPC kind even if the boundary separating the affected individuals from the unaffected was fundamentally ambiguous and the affected individuals fell on a gradient of prototypicality (Haslam 2014, p. 18). Partly for this reason, philosophers of psychiatry increasingly endorse Boyd's as the appropriate concept of kindhood for psychiatric categories (Beebee and Sabbarton-Leary 2010; Kendler et al. 2011).

Along these lines, Samuels (2009) provides the first in-depth discussion of the natural kind status of delusion in particular. He argues for the view that delusion is a natural kind in the liberal HPC sense by skillfully answering various objections to this view and drawing positive morals from them. These objections focus on three characteristics of delusion that may

be viewed as flying in the face of its natural kind status: the alleged continuity of delusion with normal experience (van Os et al. 2009); the causal, neural, and cognitive heterogeneity of delusion (Freeman and Garety 2006); and the mind-dependence of delusion as a kind (Murphy 2006). As I am confident that Samuels successfully deals with the first two groups of objections, I will not go over these here, but will confine myself to the mind-dependence objections which, I think, merit further discussion. In the next section, I will set the stage for the discussion of the mind-dependence objections by presenting a model of our intuitive detection and attribution of mental disorder, and an extension of this model that aims at accounting for the detection and attribution of delusion in particular.

2.2 Folk psychiatry and folk epistemology

2.2.1 The detection and attribution of mental disorder

How do people detect and attribute mental disorder? How do culture-specific models of dysfunction influence these processes? And how do pan-specific features of human minds influence cultural models of detection and attribution? As Pascal Boyer (2011) notes, the actual cognitive processes engaged in when people think about mental disorder have eluded empirical research. He attributes this to the fact that such processes fall between the domains of two well-established disciplines, namely, cross-cultural psychiatry (which focuses on the cultural variation of disorders themselves) and anthropological ethnopsychiatry (which focuses on cultural models of sanity and madness). Recently, however, Haslam and colleagues have, in a series of theoretical and empirical papers, developed a social–cognitive model of laypeople's thinking about mental disorder—what they dub *folk psychiatry*—which shows promise as an organizing framework for a field that has lacked a clear theoretical basis.

Haslam's folk psychiatry model specifies four dimensions along which laypeople conceptualize mental disorders: *pathologizing*, that is, the extent to which the observed behavior is construed as abnormal or deviant, mainly on the basis of rarity, and as a result of the failure to explain the behavior; *moralizing*, the extent to which the observed behavior is under the subject's control and to which individuals are morally accountable for their abnormality; *medicalizing*, the extent to which the observed behavior has a somatic basis and is the direct result of an underlying organic condition; and *psychologizing*, the extent to which the observed behavior has a mental, non-intentional basis, and is the direct result of a psychological dysfunction which shifts the explanatory focus toward causes, not reasons, undermining moral

judgment (Haslam, 2003, 2005; Haslam, Ban, and Kaufmann 2007).

Empirical support for the folk psychiatry model comes from a series of studies in which participants rate descriptions of mental disorders and other conditions on a number of items that assess features of the model. In the first study of this sort, Nick Haslam and Cezar Giosan (2002) interviewed American undergraduates who had no formal education in abnormal psychology. They were given the task of reading paragraph-length descriptions of 68 conditions, 47 of which corresponded to DSM-IV mental disorders. They were then asked to judge if the conditions were mental disorders and to rate them on 15 items addressing components of the concept of mental disorder proposed by several theorists. The authors found that American lay understandings understandings of 'mental disorder' showed moderate convergence with the DSM-IV concept of mental disorder. Then, in a follow-up study, Cesar Giosan, Viviane Glovsky, and Nick Haslam (2001) replicated the pilot study in student samples from Brazil and Romania using an identical research design and carefully translated versions of the original questionnaire. The most interesting departure from the American understanding of mental disorder was found among Brazilian participants, who did not represent moralizing and medicalizing as polar opposites, placing them on separate factors and thereby justifying the distinctness and irreducibility of these dimensions.

Besides mapping stable understandings of abnormality within and across cultures, the folk psychiatry model also illuminates shifts in these understandings. Since they found earlier that North American understandings of mental disorders tend to be more psychologized or "internalistic" than those of Brazilians, Glovsky and Haslam (2003) predicted that the longer the period of acculturation of Brazilian citizens living in the United States, the more psychologized their understandings of disorders would be compared to their less acculturated compatriots. Consistent with this prediction, more acculturated participants judged a larger proportion of the conditions to be mental disorders. Importantly, they also understood these conditions more as manifestations of emotional distress and intrapsychic dysfunction and showed a stronger tendency both to understand disorder as a violation of social expectations and to pathologize behavior in excess ('acting out'). Therefore, the concept of 'distúrbio mental' they once shared with their Brazilian peers broadened and took on a more psychologizing cast among more "Americanized" Brazilian participants.

Note, however, that while these studies and the theoretical framework that emerges from them provide an elegant illustration of the cognitive processes of intuitive detection at work, they do not address the equally important *why* and *how* questions about our intuitive detection of mental disorder—namely, why and how intuitive folk psychiatries emerge. Toward

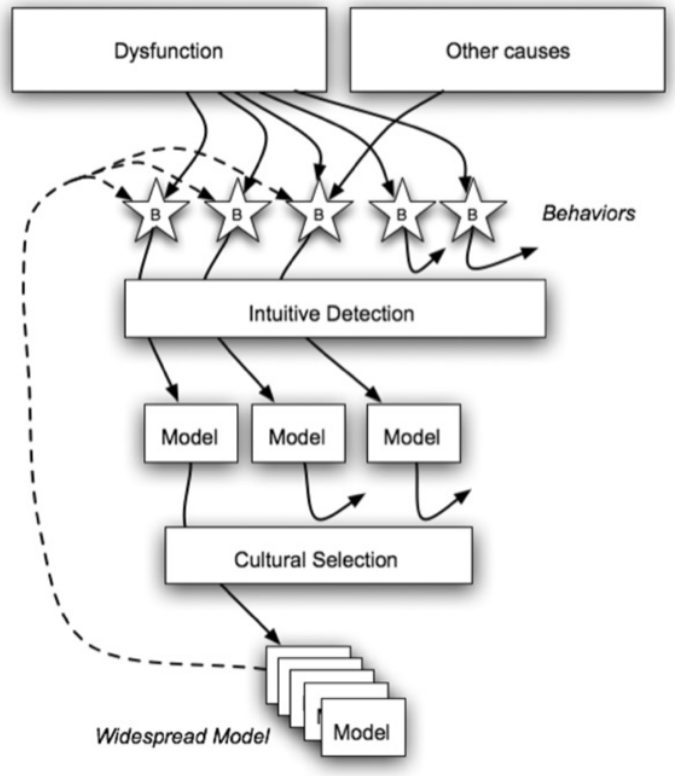

Figure 2.1: A simplified account of cognition of mental disorder (Boyer 2011)

that end, Boyer forges a cognitive model that builds on the evidence provided by Haslam and colleagues, as well as on observations about the causal connections between pathology, cultural context, typical manifestations, popular categorization, and scholarly description (see Figure 1). In the first stage of Boyer's account, dysfunction triggers behaviors, only some of which are detectable as violations of folk psychology—that is, the shared set of assumptions that are the basis of our ability to describe, interpret, and predict each other's behavior by attributing beliefs, desires, hopes, feelings, and other familiar mental states. (The ones that are not bounce off intuitive detection.) Importantly, sometimes causes other than dysfunction will trigger behaviors that will be interpreted as violations caused by dysfunction, and in these instances detection will have gone wrong. Detection of unexpected behavior will trigger explanatory causal models for the behavior, not all of which make it through cycles of acquisition and communication (unsuccessful models bounce off transmission). Finally, frequently activated models may have feedback effects. These affect the models themselves through the work

of transmission biases whereby people are more likely to adopt and transmit representations that are already widespread (Boyd and Richerson 1985). Moreover, they affect people's behaviors when subjects of classification become aware of being so classified. Such changes, in turn, may lead to revisions in the initial descriptions of mental disorders (Hacking 1995).

For our purposes, what is especially important are the first stages in Boyer's account, which, in short, boil down to the claim that our intuitive detection of mental disorder involves judging that certain kinds of behavior are so different from our expectations that they are taken as evidence that the mental systems that produce them are dysfunctional. These are mental dispositions that form part of our shared cognitive architecture (Sperber 1996). But just as 'narratives, scholarship, etiquette, politics, cuisine, musical traditions or religious rituals' (2011, pp. 112) are culture-specific, the manifestations of these dispositions to attribute dysfunction will—by deriving from the sets of mental representations that constitute the models of what is wrong with people's behavior within specific contexts—also be culture-specific. While Boyer's theory is not a theory of mental illness, but a theory of its attribution, his idea of mental disorder as a defeater of folk psychology may have an important impact on the project of uncovering natural psychiatric kinds, including the project of vindicating the natural kind status of delusion.

2.2.2 The folk epistemology of delusion

In the context of a discussion about what he calls the 'counterintuitive biology' inherent in some religious and magical concepts, Boyer (2001) considers Wendy James's account of 'ebony divination,' a practice of the Uduk-speaking peoples that she encountered while carrying out fieldwork in the borderlands of Sudan's frontier with Ethiopia in the 1960s. The Uduk report that ebony trees can eavesdrop on conversations and that they 'know of the actions of the *arum* [souls, spirits, including people who were not given a proper burial] and of *dhatu* (witches) and other sources of psychic activity' (James 1988, p. 303). According to James, diviners perform oracular consultation by burning ebony wood as a form of seeking personal healing and keeping foreign gods at bay. During the consultation, the ebony stick will produce specific smudges in the water which indicate not only the nature of the problem at hand but also a solution.

In contrast, consider the following case described by Murphy:

> Ed was sleeping rough, and heard (or, had the experience of) a tree in a park tell him that the park was a good place to stay. So

> Ed settled down for the night in the park. But a little later, the sprinklers in the park erupted and Ed was drenched. Thereupon Ed heard the tree tell him that he (the tree) was very sorry: trees like to be watered, and the tree had not understood that Ed would not appreciate a good soaking. Ed accepted the tree's apology and went on his way. (2013, p. 118)

Why is it intuitive to attribute dysfunction in Ed's case, but not in the Uduk's case? In addition to characterizing delusion as a false belief based on incorrect inference that is firmly held despite what almost everyone else believes and despite being confronted by evidence to the contrary, the DSM's definition continues in the following way: 'The belief is not ordinarily accepted by other members of the person's culture or subculture (i.e., it is not an article of religious faith)' (American Psychiatric Association 2013, p. 819). At first glance, this cultural exemption clause may appear to be a highly arbitrary, relativistic, and even unscientific addition. As epistemology does not generally regard widespread cultural endorsement as a form of justification, this sort of exceptionalism has often been dismissed as unwarranted and question-begging (Radden 2011, p. 101).

But the cultural exemption clause in the definition of delusion encodes the fact that *other causes* (see Figure 1) would be assumed rather than dysfunction in the latter case. Uduk people who believe that trees can hear conversations are members of a culture wherein trees are believed to have counterintuitive biological characteristics, whereas Ed is not. According to Samuels's interpretation of cultural exemption, in the case of the Uduk the causes of what might seem aberrant behavior for outsiders will, on close inspection, have to do with testimony: when we acknowledge that the belief that trees have counterintuitive biological characteristics is part of the Uduk culture and is acquired through testimony, the need to attribute dysfunction vanishes. In short, testimony *explains* the acquisition of strange beliefs. But what about Ed's case? Should we conversely interpret the intuitive pull to attribute dysfunction to him as being a result of Ed's *not* having the epistemic warrant that the Uduk have through testimony? As much as Samuels's observations about testimony make sense of cultural exemption in the detection and attribution of mental disorder, the converse interpretation in Ed's case makes the treatment of delusions implausible, as lack of testimonial warrant is too narrow a rationale to account for our intuitive attribution of delusion. For this reason, Murphy (2014, p. 114–5) argues that to explain the attribution of delusion we should think more broadly about reasoning, going beyond testimony.

In consonance with Boyer's cognitive account of detection and attribu-

tion, Ed's traffic with trees is readily taken as evidence of mental dysfunction in the absence of cultural exemption. Notwithstanding the fact that the description of Ed's experience is one of hallucination, the fact that he accepts this experience as true, inferring that trees can talk and letting his behavior be guided by this conviction, supports the attribution of an accompanying delusion. Murphy (2012, 2013, 2014) applies Boyer's framework to the case of delusion by hypothesizing that the psychiatric concept of delusion grows out of a widespread human tendency, which Boyer accounts for via cognitive science, to attribute mental disorder in cases where someone's behavior fails to accord with folk-psychological assumptions about how the mind works. More specifically, Murphy proposes that our practices of attribution suggest that a delusion is a belief that is acquired through a process that does not fit our folk theories of belief acquisition—which he dubs *folk epistemology*. Unlike the DSM definition, then, Murphy suggests that what is crucial to demarcating delusion from other kinds of aberrant beliefs is not the end product of reasoning but the process by which these beliefs are formed.

> What is conceptually basic about delusion is the perversion of normal mechanisms of belief acquisition and revision, not just the weird beliefs that one ends up with through that perverted changing of one's mind. "Normal" here does not mean "according to our best scientific theory." It means that folk psychology, broadly construed, endorses some avenues of belief formation and rejects others. Delusional people are people who are hooked up to the world in ways that ... folk epistemology says are weird, in the sense of falling outside normal human expectations about other people's psychology. The weirdness of the ensuing belief is (defeasible) evidence for the abnormality of their reasoning mechanisms, but the weirdness itself is not the conceptually crucial element. (2014, p. 115)

Thus, what makes delusions distinctive is not that they violate epistemic norms, per se. Instead, our folk-epistemological *expectations* are violated. All manner of beliefs that violate epistemic norms are part of our folk-epistemological expectations and can be accounted for by our folk-epistemological resources which, Murphy (2012, p. 22) elucidates, do not just include folk psychology in the narrow sense of theory of mind, but also beliefs and expectations about the role of "hot" cognition and personal interests in the formation and maintenance of belief, as well as the role of culture in shaping people's assumptions about what counts as legitimate evidence. In the case of self-deception, for example, though the belief is formed and maintained in the face of contradictory evidence, we as interpreters do not run out of

explanatory resources and can readily come up with an explanation of how and why the belief came about. In other words, what is distinctive about delusion is the "explanatory gap" created by its observation, and closed by its attribution.

2.3 Assessing the mind-dependence of delusion

How does Murphy's Boyer-inspired account of delusion attribution impact the status of delusion as a natural kind? Unlike biological taxa which, as we have seen, are prime examples of property clusters held together by homeostatic causal mechanisms, delusion (as well as other psychiatric categories) appear to be mind-dependent (or response-dependent) in ways that put pressure on even the most liberal sense of natural-kindhood.

2.3.1 Delusion as a folk-psychological kind

The first mind-dependence objection one may extract from the discussion of the attribution of delusion simply states that delusion is not a natural kind because it is an artifact of our folk psychology. As Murphy claims, 'whether or not something is a delusion is a matter of how it strikes us, and that depends on how well it comports with our understanding of what people are like, both in general terms and within our culture' (2006, p. 180). Note, however, that even if we follow Samuels and derive such an objection from Murphy's claim that delusion is a matter of how it strikes us, this objection could not be derived from the mere fact that delusions are a part of our folk conception of the world, since there is no immediate incompatibility between the naturalness of a kind and the fact that it maps onto our folk conceptions.

As Samuels notes, water is plausibly a natural kind, though 'water' and the concept it expresses are also part of our folk conceptions. Though one may have affinities for eliminativism concerning some of our folk concepts, there is, on the other hand, no principled reason to deny that at least some of our folk concepts do pick out natural kinds. What the present objection hinges on is the premise, attributed by Samuels to Murphy, that *what it is to be a delusion* is determined by how it strikes us. That is, the premise that all there is to being a delusion is to be a certain kind of response-dependent property. As we have seen, Samuels alludes to Page (2006)'s notion of individuative independence—the sense in which a class of things is circumscribed by boundaries that are totally independent of our taxonomic practices—as the relevant sense in which natural kinds must be response-independent. So

the objection at hand can be seen as likening the individuation of abnormal psychological conditions into delusions to the individuation of the night sky into constellations: just as the existence of constellations is parasitic on the way we choose to categorize things, so does the existence of delusions. In other words, the task for those who wish to argue that delusion is a natural kind consists in showing that delusion as a kind is more akin to stars than to constellations.

Samuels's answer to the response-dependence objection consists in arguing that it conflates the metaphysics of delusion with its epistemology: 'The relevant metaphysical issue concerns the nature of delusions: roughly, what is it to be a delusion. The relevant epistemic question concerns the evidential basis for our judgements about delusion: roughly, the sorts of evidence we invoke in judging that someone is deluded' (2009, p. 68–69). Samuels concedes that Murphy gets the epistemology of delusion right, and that not only everyday judgments about which mental states are delusions are made on the basis of commonsense psychological considerations, but the judgements of clinicians who diagnose delusions are also largely dependent on the same folk conceptions. Samuels's point, then, is that the fact that the detection and attribution of delusion is a matter of how it strikes us does not show that what it is to be a delusion is *exhausted* by how things strike us and, consequently, there is still a possibility that, in this case, our folk conception will be vindicated by, and map onto, a scientific understanding of delusion—what Murphy (2014, p. 119) aptly calls the *vindication project*.

2.3.2 The cultural relativity of delusion

The second mind-dependence objection to which Samuels refers is that which states that delusion is not a natural kind because delusion is context-sensitive. In fact, there are two senses in which delusion may be said to be culturally relative. The first sense expands on what has been just discussed, namely, the fact that the *attribution* of delusion derives from our folk conception of what is and isn't a healthy or normal state of mind. Whereas the previous objection concerns an allegedly universal feature of human folk psychology, a new objection may hinge on the claim that the attribution of delusion will also depend on what is considered a healthy or normal state of mind within one's cultural context, encoded in the cultural exemption clause in the definition of delusion given in the DSM-5. The clause makes sense of the intuition that the delusional individual *stands alone* in some sense (Leeser and O'Donohue 1999, p. 692). The intuitive character of the cultural exceptionalism clause can be seen by contemplating what we would judge as strange and even irrational beliefs which are nevertheless commonplace in cultures other than our own.

For example, consider the following entry in Dan Sperber's field diary, from the period he conducted ethnographic fieldwork among the Dorze people of Southern Ethiopia between 1968 and 1974:

> Saturday morning old Filate came to see me in a state of great excitement: "Three times I came to see you, and you weren't there!"
> "I was away in Konso."
> "I know. I was angry. I was glad. Do you want to do something?"
> "What?"
> "Keep quiet! If you do it, God will be pleased, the Government will be pleased. So?"
> "Well, if it is a good thing and if I can do it, I shall do it."
> "I have talked to no one about it: will you kill it?"
> "Kill? Kill what?"
> "Its heart is made of gold, it has one horn on the nape of its neck. It is golden all over. It does not live far, two days' walk at most. If you kill it, you will become a great man!"
> And so on . . . It turns out Filate wants me to kill a dragon. He is to come back this afternoon with someone who has seen it, and they will tell me more . . . (1982, p. 35)

Commenting on this entry, Sperber goes on to express respect and affection for his Ethiopian friend. He is confident that the man was not senile at the time of the unusual request and, moreover, that he was too poor to drink. Consequently, Sperber is faced with a variation of a question that, undoubtedly, all of us ask ourselves of someone else at some point: how could a sound person believe *that*? 'That' being, in this case, that dragons exist, not "once upon a time," but there and then, within walking distance. What if Sperber had expressed doubts that such an animal even exists? What if he had pressed his friend on the issue of the dragon's heart being made of gold and the apparent impossibility of a gold heart *beating*? Sperber concludes that his friend was 'merely quoting what people who had killed these animals were reported to have said, and they knew better than any of us' (1982, p. 61). In line with Sperber's explanation, Samuels (2009, pp. 69–70) argues that the cultural relativity of delusions tracks precisely the insensitivity of delusions to testimony—an important source of epistemic warrant and epistemic defeat. Because it is normal for one to form and maintain beliefs based on the testimony of peers and authorities from one's culture or subculture, resistance to testimony is viewed as a sign that something is wrong. And because one's source of testimony varies with one's culture and subculture, the cultural exemption clause is a necessary measure to avoid the hasty judgment that culture-bound beliefs are necessarily irrational and possibly even the

product of pre-rational mental processes (Sperber 1980). However, so long as the resistance to testimony that characterizes delusion is culturally *invariant*, the fact that delusions are resistant to testimony does not suffice to show that delusion is a response-dependent property and, thus, cannot be used to successfully object to the natural kind status of delusion.

The second sense in which delusion may be said to be culturally relative derives from the fact that the *content* of delusions is highly sensitive to social and cultural context. So, for example, Masato Tateyama and colleagues (1998) compared the schizophrenic delusions of 324 inpatients in Japan, 101 in Austria, and 150 in Germany, and found that themes of persecutory delusion (i.e., delusions of poisoning) and religious themes of guilt/sin were conspicuous in Europe, while amorphous delusions of reference (i.e. 'being slandered') were predominant in Japan. Another study conducted by Thomas Stompe and colleagues (1999) compared the schizophrenic delusions of 126 Austrian and 108 Pakistani patients, finding significantly higher frequencies of grandiose and religious delusions in Austrian patients, and persecutory delusions with political themes among male Pakistani patients. To these observations may be added the existence of culture-bound syndromes whose expression includes culture-specific symptoms, as in *koro*, most prevalent among Chinese ethnic groups, in which an individual claims that his or her genitals are retracting and will disappear (Chowdhury 1996).

Time is also a factor. Changes within one and the same culture have an impact on the diachronic variability of delusional content, as Borut Škodlar and colleagues (2008) have found in a study of admission records of patients with schizophrenia in Slovenia from 1881 to 2000. The recent emergence of the so-called Truman Show delusion attests to the same fact—patients with 'Truman signs' claim that their lives are staged plays or reality television shows, as with the protagonist of the 1998 film *The Truman Show* (Fusar-Poli et al. 2008; Gold and Gold 2012). However, though the kinds of variability discussed above may suggest that delusion is response-dependent to the extent that what is a delusion depends on what beliefs are socially prevalent at a certain point in time, Samuels (2009, p. 69) notes that what the sensitivity of delusions to social context shows is only that the nature of delusion, as Karl Jaspers (1913) long before observed, cannot be characterized, but can at best only be classified, in terms of its contents.[14]

2.3.3 The vindication project

If Boyer and Murphy are correct, then the science of delusion is inextricably tied with its intuitive detection. Psychiatric elaborations of folk psychology

[14] See section 1.1 of chapter 1 of this dissertation.

give rise to the clinical concept of delusion, the extension of which is then subdivided according to surface features, most prominent among these its content (i.e., what it is about). But can delusion, being rooted in folk psychology, play the role of regimenting scientific inquiry?

By defending that delusion is a natural kind in the HPC sense, Samuels answers positively and wagers that scientific psychiatry will vindicate the folk concept of delusion—that is, if Samuels is correct, the folk concept of delusion picks out a causal signature that, once uncovered, will vindicate the reliability of this concept and show that delusion is, in fact, a homeostatic property cluster. Once the causal mechanisms that make the properties of delusion co-occur are discovered, causal classification may result in many current subtypes of delusion being excluded from its extension. But because the HPC conception of natural kindhood does not mandate that natural kinds have category essences or category boundaries, it is likely that a mature science of delusion informed by its causal mechanisms will not be able to give a simple yes or no answer to every question of the form 'Is X a delusion?'. Of more practical importance, however, is the fact that a causal understanding of the underlying mechanisms would suffice to yield powerful inductive generalizations regarding diagnosis, prevention, and management.

But how does the vindication project fare in view of the mind-dependence of the folk concept of delusion? As Samuels notes, this only hurts the chances of delusion being an HPC kind if we conflate the metaphysics and epistemology of delusion. As we have seen, Samuels argues that attention to the fact that our concept of delusion is a part of our folk psychology that has been incorporated into scientific psychology and psychiatry is not enough to show that it is *not* a natural kind: the folk-psychological kind may well track an underlying natural kind. Samuels (2009, p. 69) notes that, to support the mind-dependence objection, it would be necessary to show that in the case of delusion the metaphysical issues about the nature of the kind and the epistemic issues about how we know about instances of the kind *should* be collapsed. Showing that the clinical concept is built on folk conceptions of normality is not enough. Importantly, however, Samuels does not establish that delusion *is* a natural kind. In fact, he could not have established this on the basis of *a priori* speculation alone, as establishing natural kindhood is ultimately a matter of investigating the causal basis of the homeostasis of property clusters (assuming the HPC model). Samuels does skillfully argue against various objections to the status of delusion as a homeostatic property cluster, some of which I have discussed above. In doing so, Samuels establishes something very important, namely, that these objections are not sufficient to exclude the possibility that delusion is a natural kind. So what we are left with after Samuels's arguments is that the natural kind status of

delusion is still an open question, i.e. that delusion is *possibly* a natural kind.

Although the argument from mind-dependence that derives from accepting the application of Boyer's theory to delusion is not enough to rule out the possibility that delusion is a natural kind, it does make Samuels's thesis implausible and gives him the burden of proof. This implausibility can be better seen if we compare generic folk kinds and generic scientific kinds. If Samuels is right, delusion would be a *generic* natural kind. Just like the kind metal subsumes many different subordinate kinds such as gold, copper, and magnesium, delusion will subsume subtypes which would themselves also be natural kinds. But Samuels's optimism regarding the vindication project is hardly justified by the observation of other generic folk concepts and how they relate to their scientific counterparts. For instance, what the folk concept of metal seemingly picks out is not a causal signature, but, as Murphy (2014, p. 121) notes, a variety of properties that directly relate to our interests, properties like being shiny, being malleable, etc., rather than a chemical element whose atoms readily lose electrons to form positive ions, etc. Likewise, the folk concept of lily, as Dupré (1981, p. 74) points out, does not accurately map onto the biological concept of lily, which includes garlics and onions, but is used to refer exclusively to a type of flower. If delusion picks out properties that relate to our interests, like being *weird* to varying degrees, then the burden of proof falls squarely on Samuels with respect to the likelihood of vindication.

Furthermore, as investigation into the causes of delusion is still in early stages, accepting the view that delusion constitutes an HPC kind is as much a "black-box" approach as Wakefield's, only more modest in its ambition. I have argued that as an ontological commitment, this approach is weak. As a *methodological* commitment, on the other hand—and this is the sense in which Kendler and colleagues (2011) seem to accept that psychiatric categories in general are HPC kinds—there is still a case for viewing delusion as a generic natural kind with an eye toward progress in scientific psychiatry. Bearing in mind that what we are authorized to commit to (ontologically) at this moment is that delusion is a *practical* kind—as this coheres both with our knowledge of delusion in the clinic as well as with our best theory of detection and attribution of mental disorder (and delusion in particular)—if the possibility of natural kindhood is still open, assuming natural kindhood is a sound methodology inasmuch as it offers a way toward progress in causal classification. However, I maintain that this is neither the only, nor the best way toward progress.

Even if the *folk* concept of metal is not appropriate to play the role of regimenting scientific inquiry, chemistry did eventually arrive at the natural kind metal and many subspecies of our folk concept of metal, such as gold,

silver, copper, etc. also turned out to be natural kinds. In this manner, despite delusion being a folk concept not so far mapped onto a rigorous scientific concept, many subtypes of delusion already recognized, such as clear-cut cases of monothematic delusions following brain damage (e.g. Capgras, mirrored-self misidentification, somatoparaphrenia, etc.), might still turn out to be natural kinds which are thrown in with similar conditions that strike us as weird into the set of phenomena described folk-psychologically (and clinically) as delusions. Our focus should be on uncovering the causal mechanisms underlying specific kinds of delusion rather than trying to impose a general causal explanation on a ragbag of different abnormalities that may or may not actually be of the same kind. Thus, I suggest a compromise between Zachar's (2000) earlier work and Samuels's (2009) defense of delusion as an HPC kind, drawing on Murphy's (2014) observations: delusion, as a kind rooted in folk psychology, is probably a practical kind, and it probably does not pick out a universal causal signature that makes the whole category be a natural kind, but it probably *does* pick out many subspecies which are themselves natural kinds. Hence:

> *Hypothesis:* Delusion is a not a natural kind, but some delusions are.

So if the question were 'Is Capgras delusion a natural kind?,' or 'Is somatoparaphrenia a natural kind?,' being that these are stable clusters of properties with recognizably homogeneous neurological causes and which are not the product of generic folk intuitions but of rigorous clinical observation and investigation, the case for their natural kindhood would be much stronger and plausible. Thus, I suggest that the way to progress in the science of delusion lies in trying to vindicate the natural kind status of *subspecies* of delusions through the study of the causal mechanisms that make the relevant properties occur homeostatically, and not in trying to find a shared causal basis for every phenomena that *we* call delusion assuming beforehand that such a share causal basis is present. After the investigation into the causal mechanisms is done with multiple subtypes of delusion, a causal account of delusion in general will no doubt progressively suggest itself. But the set of delusion subtypes that will be found to share causal mechanisms in the sense that would authorize us to abstract from them a generic natural kind will be a *subset* of the set of all delusions—a set the intension of which depends on context-dependent folk-psychological intuitions and, hence, membership in such a set is tied to surface features (symptoms, not causes) detected with the tools of folk psychology.

Conclusion

In the preceding sections, I have attempted to elucidate some of the difficulties inherent in trying to claim that delusion is a natural kind. After delineating five different senses of kindhood and introducing a non-essentialist approach to natural kindhood—the HPC model—I have drawn on a cognitive model of the intuitive detection and attribution of mental disorder and its application to the case of delusion to flesh out the fact that the clinical category of delusion is rooted in folk-psychological expectations. Finally, being that the folk-psychological status of delusion does not immediately remove the possibility of this kind being vindicated as natural by scientific investigation, I have questioned the vindication project and formulated a working hypothesis that I claim is both ontologically and methodologically more sound. My hypothesis is that along with the general category of delusion, some delusions will be confined to practical kindhood, perhaps along with the bulk of mental symptoms and disorders, while some will turn out to be objective distinctions in nature. Importantly, this hypothesis and methodological suggestion bypasses what Samuels calls the *unity problem*: if many different subtypes of mechanism are responsible for delusions, why treat delusions *as such* as a natural kind? According to him, it must be because these mechanisms are themselves of the same kind. What I have tried to show in this chapter is that this is an improbable scenario. Assuming that a variety of mechanisms make subtypes of delusion subtypes of some general mechanism as opposed to a heterogeneous collection of different mechanisms the products of which share surface features is not only unwarranted, but methodologically flawed.

Chapter 3

The doxastic status of delusion

Introduction

Clinical delusions are commonly thought of and characterized as *beliefs*, both by psychiatrists and by the general population. Here is the definition of delusion in the 'Glossary of Technical Terms' of the most recent edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5):

> A false belief based on incorrect inference about external reality that is firmly held despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. (American Psychiatric Association 2013, p. 819)

Although almost every aspect of this definition is debatable (Coltheart 2007), describing delusion as a type of aberrant belief is the only one to have engendered a specialized literature in itself, engaging philosophers, psychiatrists, and psychologists in the project of arriving at a precise characterization of this class of mental states. Intuitively, delusions do seem like they warrant the attribution of beliefs. This is mainly because of patients's verbal behavior, represented both in the assertions of delusional subjects and the apparent sincerity with which those assertions are made. Take Capgras syndrome, for example. Patients with Capgras are characterized by their inability to recognize a loved one, a close relative, or a friend (or sometimes multiple persons and sometimes even animals and inanimate objects). As Adriano Rodrigues and colleagues go on to explain:

> In this monothematic delusion, the individual recognizes overtly and straightforwardly who that person is meant to be, upholding

> however a firm *belief* to the contrary, which is anchored in subjective cues such as an eerie feeling that something is not quite right about that person, complete lack of a sense of familiarity, and missing the proper affective response. Individuals with Capgras syndrome cling to the unshakeable *belief* that the original person in question was replaced by an impostor, who cunningly is trying to fool them—with no success at all because, of course, they know better' (2013, p. 522, my emphases).

Recently, David Rose, Wesley Buckwalter and John Turri (2014) have presented evidence from five studies that folk psychology not only unambiguously views monothematic delusions as beliefs, but that it views delusions as *stereotypical* beliefs. Furthermore, they show that frequent assertion is a powerful cue to belief attribution, more powerful than even a robust and consistent track record of non-verbal behavior. As we will see in section 2, in the specialized literature the presence of certain kinds of non-verbal behavior is one of the main reasons pointing toward the opposite attribution (or at the very least the withholding of attribution) and supporting the abandonment of the view that delusions are beliefs (henceforth *doxasticism about delusion*). Subsequently, in section 3, I will present the main alternative characterizations that have emerged in the wake of doxasticism. Then, in the remainder of the chapter, I will question the validity of the debate between doxasticists and non-doxasticists by stepping back and assessing the meaning and relevance of the question 'Are delusions beliefs?'. I will argue in sections 4 and 5 that, by focusing on what appears to be a merely terminological dispute, the theorists engaged in this debate have lost sight of two critical aspects of a precise characterization of delusions, namely, its use in the development of a scientific theory of the relevant phenomena and its ability to account for the experience of the patients.

3.1 Problems for doxasticism

The implausibility of ascribing full-fledged belief to delusional subjects has been hinted at since at least the 1910s, when both Karl Jaspers's *General Psychopathology* and Eugen Bleuler's *Textbook of Psychiatry* were published. The set of objections against the traditional view forms an unavoidable obstacle for doxastic accounts.[1]

[1] In listing these objections I largely follow the excellent survey provided in Bayne and Pacherie (2005). I also benefited from the discussion in Stephens and Graham (2004), Egan (2009), and Bortolotti (2010).

3.1.1 Content and evidence

One objection—originally raised by Jaspers (1913/1963) and elaborated recently by German Berrios (1991) and Louis Sass (1994)—denies that delusions are contentful states. One may call this the *expressivist* (Gerrans 2001) or *non-assertoric* (Young 1999) account. This view is motivated by the fact that most (if not all) delusions appear obviously false or incoherent. Berrios, for example, states that when a patient who utters a verbal formula such as 'I am dead' or 'My internal organs have been removed' is questioned as to the real meaning of these assertions, she will not be able to coherently discuss them or their implications. 'Properly described,' says Berrios, 'delusions are *empty speech-acts* that disguise themselves as beliefs' (1996, 126, my emphasis). 'Their so-called content refers neither to world, nor self'. 'Delusions are so unlike normal beliefs that it must be asked why we persist in calling them beliefs at all' (1996, 114-5). A wide variety of other cases besides Cotard's can be summoned in favor of such a view. Tim Bayne and Elisabeth Pacherie (2005) cite an intermetamorphosis patient who claimed that his mother changed into another person every time she put her glasses on (De Pauw and Szulecka 1988); another that had the delusion that there was a nuclear power station inside his body (David 1990); and a third that had the delusion of being both in Boston and in Paris at once (Weinstein and Kahn 1955).

One may not want to deny that delusional states possess content, and still object that it is difficult to see *how* the delusional patient themselves could believe such content. Again, Cotard patients are a fitting example. José Luis Bermúdez voices this concern in stating that there is 'something content-irrational about the belief ... that one is dead—because, to put it mildly, the belief is *pragmatically self-defeating*' (2001, p. 479, my emphasis). Not only is it unclear that a self-defeating assertion such as 'I am dead' could be coherently expressed,[2] the question is open whether there can be self-defeating *beliefs* to begin with (as opposed to mere verbal utterances).

Still, one may point out that delusional subjects appear to lack reasons or evidence for their delusional state. However faulty the reasons or flimsy (and biased) the evidence one may have to support some self-deceptive belief, there will be nevertheless *some* kind of support for such a belief. In contrast with this, John Campbell cites the well-known case of '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' (2001, p. 95). Notwithstanding the DSM definition

[2] Except, of course, in such contexts as that of the opening words of a will ('Now that I am dead...'), or of the hero's epitaph in Ezra Pound's *Mauberley* ('I was. And I no more exist; Here drifted. An hedonist.').

of delusions (that they are held 'despite what constitutes incontrovertible and obvious proof or evidence to the contrary'), Campbell points out that it is difficult to understand (to put it mildly) how an experience of marble tables could verify the proposition 'The world is ending'.[3] On the other hand, there is at any time a considerable body of evidence *against* the truth of the delusional content, to which the delusional subject seems utterly impervious. Furthermore, there are delusional patients that even recognize that they do not have evidence for their claims. A case in point is Andrew's Young and Katherine Leafhead's Cotard patient, JK:

> We wanted to know whether the fact that JK had thoughts and feelings (however abnormal) struck her as being inconsistent with her belief that she was dead. We therefore asked her, during the period when she claimed to be dead, whether she could feel her heart beat, whether she could feel hot or cold. ... She said she could. We suggested that such feelings surely represented evidence that she was not dead, but alive. JK said that since she had such feelings even though she was dead, they clearly did not represent evidence she was alive. (1996, pp. 157–8)

This is a startling case of cognitive dissonance and resisting evidence, and suggests something along the lines of Andy Egan's observation that 'if we think that a certain responsiveness to evidence is essential to belief, then, in many cases, we'll be reluctant to say that delusional subjects genuinely believe the content of their delusions' (2009, p. 266). In other words, if there is a constitutive relationship between belief and evidence (even in the case of irrational belief and improper evidence), then it seems that delusional states do not warrant the ascription of delusional beliefs. This paves the way to what is perhaps the most objection to the doxastic conception: those which point to bad integration between the subject's delusion and his or her other beliefs (Bortolotti 2010).

3.1.2 Circumscription

Delusional states present a degree of *circumscription* (Young 1999, p. 581) that may speak against their being properly taken as beliefs. Egan calls this

[3] Although not impossible. Paulo Faria (personal correspondence) suggests the following scenario: 'Suppose the Almighty (under cover, perhaps, of a burning bush, as in Exodus 3, 2-21) had told his prophet, call him Moses II, that the end was approaching, and that, as a warning signal to His chosen children, he would have Moses II run against a row of empty marble tables when entering a café. (We may suppose it would then be Moses II's duty to warn his brethren that the end had come.)'.

property of delusional states *inferential* circumscription (2009, p. 266). As Bayne and Pacherie neatly put it:

> A subject will normally accept the obvious logical implications of her beliefs—at least when these are pointed out to her. And when she realizes that some of her beliefs are inconsistent, she will normally engage in a process of revision to restore consistency. In contrast, deluded patients often fail to draw the obvious logical consequences of their delusions and show little interest in resolving apparent contradictions between their delusion and the rest of their beliefs. (2005, p. 164)

This is the precisely the vein in which Currie and Ravenscroft affirm that

> If someone says that he has discovered a kind of belief that is peculiar in that there is no obligation to resolve or even to be concerned about inconsistencies between these beliefs and beliefs of any other kind, then the correct response to him is to say that he is talking about something other than belief. (2002, p. 176)

However, the majority of patients with the Capgras delusion, for example, do not draw the consequences the content of their delusion would usually mandate: their *worldview* does not seem to change at all as a consequence of supposedly adopting the belief that their spouses have been abducted and that the person they see in front of them is an impostor (Davies and Coltheart 2000). Whatever this state is, therefore, it seems that it is severely encapsulated, failing to be integrated with the subject's web of belief. But beliefs are the mainstay of theoretical and practical reasoning and, while one may ascribe false belief to subjects for any number of reasons, a state that fails to have the appropriate connections to the subject's other mental states may not be properly described as a belief.[4] As exemplified by Currie and collaborators, this view is especially espoused by authors who (tacitly or explicitly) endorse a consistency constraint on belief-ascription.

Indeed, authors such as Quine and Ullian (1970), as well as Fodor (1983), have argued that one of the attributes of a belief *qua* belief is its property of being inseparably connected with other beliefs of potentially widely diverse contents. Quine's answer as to why beliefs should be webbed or interconnected with other beliefs in a way that precludes severe encapsulation rests on the conditions of epistemic assessment of beliefs—for instance, whether

[4] I say 'appropriate' rather than 'necessary' because circumscribed (insulated) "beliefs" will usually stand in a number of (nonlogical) connections to the subject's other mental states: that of being simultaneously held to begin with, and then that of causing or being caused by other mental states.

I am warranted in believing that an acquaintance of mine lives in Chicago may depend on whether I believe that Chicago is a city and believe that cities are bigger than towns, etc. And for Quine, the conditions of epistemic assessment of beliefs are part of their functional role: beliefs are states or attitudes that are constituents in (what Fodor calls) the central processing that takes place in the mind.[5] Therefore, like-minded theorists will deny that delusional subjects are in the hold of belief.

Belief has important connections to action, and many delusional subjects fail to act in ways expected of agents who really believed the content of their delusions. As Currie puts it, delusion 'exerts a powerful psychological force, absorbing inner mental resources, but it fails to engage behavior in the way that genuine belief would' (2000, p. 175).[6] This seems likely due to the inferential circumscription noted above. Egan calls this characteristic of delusional patients *behavioral* circumscription (2009, p. 266). It was noted by Bleuler, who stated that his delusional patients 'rarely follow up the logic to act accordingly, as, for instance, to bark like a dog when they profess to be a dog. Although they may refuse to admit the truth, they behave as if the expression is only to be taken symbolically' (1916/1924). In the same manner, Capgras patients who (for all we can see) sincerely affirm 'This is not my wife' or 'My mother has been replaced by an impostor' do not as a consequence of this go looking for their missing loved ones, nor do they call the police to report the breaking and entering perpetrated by the person they claim to be an impostor.

Finally, delusional patients often fail to exhibit the affective (i.e. emotional) responses one would expect of a person who believes the content of her assertions (Sass 1994, pp. 23–24). We may call this *affective* circumscription, since what is observed is a failure of integration between the subject's delusional state and their emotional lives. Capgras patients are more often than not unmoved by the fate of their relatives whom, according to the doxastic interpretation of this delusion, they believe to have been abducted. Why don't they exhibit the affective responses which the relevant beliefs would

[5] Fodor is committed to the analogy between scientific confirmation and psychological fixation of belief, and states that 'the central processes which mediate the fixation of belief are typically processes of rational nondemonstrative inference and that, since processes of rational nondemonstrative inference are Quineian [i.e. the degree of confirmation assigned to any given hypothesis is sensitive to properties of the entire belief system] and isotropic [i.e. the facts relevant to the confirmation of a scientific hypothesis may be drawn from anywhere in the field of previously established truths], so too are central processes. In particular, the theory of such processes must be consonant with the principle that the level of acceptance of any belief is sensitive to the level of acceptance of any other and to global properties of the field of beliefs taken collectively' (1983, p. 110). [6] See also Sass (1994, p. 21) and Young (1999, p. 581).

lead us to expect?

Bortolotti observes that 'although it is possible for a belief system to have some internal tension, most philosophers resist the thought that subjects capable of having beliefs can have dissonant attitudes simultaneously activated and operative at the forefront of their minds' (2010, p. 62). Delusions lack the holistic character expected of beliefs and do not respect the notion of a coherent belief system whose adjustments to one belief implies adjustments to many others (Young 2000, p. 49). Belief-ascription in the context of delusion, then, is only admissible after explaining away these disparities between the roles that delusional states play in the overall cognitive economy of delusional patients and those roles we expect beliefs to play (following either folk-psychological intuitions or fully articulated theories of belief).

Why would these features put pressure on the thesis that delusions are beliefs? Inherent in the notion that they work as an objection to doxasticism is a hidden premise, namely, a vague functionalism about belief. Acknowledging this suffices for us to extract from the literature the following general argument against doxasticism:

1. Playing a belief-like functional role is necessary for a mental state to be a belief.

2. Delusions fail to play belief-like functional roles.

3. Therefore, delusions are not beliefs.

As we will see in the next section, many philosophers in the literature accept this argument and reject doxasticism, while other philosophers resist the argument by rejecting its second premise.[7] Those that accept the argument from functional role have suggest alternative, *non-doxasticist* theories of delusion, to which I now turn.

3.2 Alternative attitudes

In response to the problems raised in the last section, mainly two kinds of alternative accounts have emerged. Some authors have tried to characterize delusions with the resources of traditional folk-psychology, and they have done so either by insisting that while delusions are not beliefs, their status can be captured by other familiar kinds of propositional attitudes (e.g., imagination). Other authors instead propose a revision to standard folk-psychological categories, claiming that delusions are a hybrid type of propositional attitude

[7] The argument's first premise has gone mostly unnoticed. But see Miyazono and Bortolotti (forthcoming).

that was simply not recognized before (e.g., "bimagination"). Finally, a third kind of response consists in taking the difficulties of attribution presented by delusion as forcing us to rethink drastically the nature of belief. In this section, I will sketch the first two kinds of response, presenting a few problems for these approaches in section §4, and dealing with the third kind of response in section §5.

3.2.1 Imagination

Gregory Currie and colleagues (Currie 2000; Currie and Jureidini 2001; Currie and Ravenscroft 2002) argue that delusions—or at least some delusions, especially those manifested in schizophrenia[8]—are not straightforward, first-order beliefs, but rather *cognitive hallucinations*: imaginative states that are misidentified by their subjects as beliefs. As Currie puts it, 'what we normally describe as the delusional belief that p ought sometimes to be described as the delusional belief that I believe that p' (2000, p. 175). Because Currie's account involves reference to deficits in the self-monitoring of mental states it has been referred to as the *metacognitive* or *metarepresentational* account of delusion. Bayne and Pacherie (2005, pp. 165–166) identify three claims in Currie's account (where p is the content of the delusional state):

1. Delusional patients who seem to believe P do not actually believe P;

2. Delusional patients who seem to believe P actually imagine P;

3. Delusional patients who seem to believe P believe that they believe P.

There are at least two important aspects to Currie's account that deserve attention. First, it is anchored in an influential and powerful theoretical model of schizophrenia. In his book *The Cognitive Neuropsychology of Schizophrenia*, Chris Frith (1992) argues that various symptoms of schizophrenia are the result of an underlying deficit of metarepresentation, the capacity to formulate thoughts about thoughts. Currie's description of thought monitoring and his explanation of how the delusional imagining is mistaken for a belief derives from Frith's model of disorders of volition, such as passivity phenomena and alien hand syndrome. Frith sets out to explain why schizophrenics often

[8] The intended scope of Currie and colleagues's accounts is not completely clear. As Bayne and Pacherie (2005, p. 166) note, they apply it to the florid and polythematic delusions typical of schizophrenia, but they are less explicit about whether or not it applies to monothematic delusions. In some places Currie implies that the model should be extended to include monothematic delusions (Currie and Jureidini 2001), while in other places he suggests that a different account of monothematic delusions might be appropriate (Currie, 2000).

claim that their bodies are being controlled by someone else, or that their thoughts are not their own, and in answering this question develops what is called the *efference copy* model of volitional action. An efference copy is an internal copy of an outflowing, movement-producing signal generated by the motor system (Jeannerod 2003, p. 83). Once a person forms an intention to act in a certain way, a command is sent to motor control. When a motor instruction is sent for bodily movement, a copy of that instruction—the 'efference copy'—is also sent to some other center. John Campbell provides an excellent explanation of the ensuing model:

> [Richard] Held (1961) suggested that copies of the motor instruction are sent to a comparator, stored there, and compared to the proprioceptive or visual—'reafferent'—information about what movement was actually made. . . . What explains the feeling that it is you who moved your arm is that at the comparator, an efferent copy was received of the instruction to move your arm which matches the movement you perceive. What explains the feeling that your arm was passively moved, perhaps by someone else, is that there is no efferent copy at the comparator of an instruction to move the arm in a way that matches the movement you perceive. (1999, pp. 611–12)

Hence, if there is no efferent copy with matching content, then the movement is experienced as not controlled by the agent and, therefore, non-volitional. This explains why we normally experience a felt difference when we raise our arm and when our arm is raised for us. In some schizophrenic patients (e.g., those suffering from an alien hand), however, Frith proposes that there is impairment of action monitoring in schizophrenia due to impaired efference copying (a "broken" comparator), and that there is comparably based impairment to intention monitoring. Thus it becomes difficult for the schizophrenic person to detect her own actions, and also her own acts of will.

Note, however, that Frith's postulation is that of a failure in a *subpersonal* mechanism, rather than the postulation of a difficulty with the personal-level mechanism of metarepresentation. But if Frith is right concerning patients that suffer from an alien hand, what about those that suffer from, say, delusions of thought insertion? Approvingly citing the idea that thinking is a kind of motor action, Currie's model extends Frith's thesis about impaired self-monitoring of action to the self-monitoring of *thoughts*: some schizophrenic patients don't feel that their thoughts are their own because of a failure in efference copying. Richard Dub aptly explains how this connects the failure to recognize one's own thoughts as self-generated to the imagination:

> One of the features of imagination, according to Currie, is that imaginings are normally *recognized* as such by their subjection to the will. This is a variant on the Wittgensteinian notion that imagination is distinguished from perception by being subject to the will. Currie's amendment is that imagination is recognized as imagination by being *felt* as if willed. Because of a broken comparator, the schizophrenic does not take an imagining that p to be his, and so does not recognize it as an imagining that p. Instead, the free-floating thought that p is experienced as a belief. So, the schizophrenic comes to believe he believes that p. (2013, p. 68)

But note that a problem for Currie's theory surfaces with the fact that it is not altogether clear to which attitude we should affix the term 'delusion': whether we should say that the subject's imagining that p or her believing that she believes that p is the delusional state. On the one hand, if we are to take Currie's theory to resolve the problems left by doxasticism, then we should understand delusions to be imaginings. After all, the second important aspect of Currie's model—especially relevant for the purposes of the present discussion— is that it promises to account for the features of delusion that are not well accounted by the doxastic account: '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 undisturbed by such inconsistency, and to be immune to conventional appeals to reason and evidence' (Currie and Ravenscroft 2002, p. 179). Furthermore, Currie claims, his model can account for the fact that delusions typically fail to result in direct actions or strong affective responses, since this is also true of imaginings.[9]

However, the way his model is presented, and the very terminology of 'cognitive hallucinations', undeniably suggest that delusions are not the imaginings, but rather the second-order (metarepresentational) beliefs which are themselves *caused* by wayward imaginings (Dub 2013, p. 70). Indeed, simply *imagining* that one is dead, or that one's spouse has been replaced by a double, or even that that divine forces were preparing one for a sexual union with God (Schreber 1903) should certainly not be seen as tantamount to being delusional. The wavering equivocation in Currie's account is made especially clear in the following passage, where he and Nicholas Jones have more recently suggested, tentatively, that delusions considered as a class of states do not fit easily into rigid categories of either belief or imagination.

[9] On the properly metarepresentational portion of Currie's model, it also promises to explain the patient's verbal behavior: the patient says (that she believes that) p because she believes that she believes that p.

> While delusions generally have a significant power to command attention and generate affect, they vary a great deal in the extent to which they are acted upon and given credence by their possessors. In that case it may be that cognitive states do not sort themselves neatly into categorically distinct classes we should label 'beliefs' and 'imaginings', but that these categories represent vague clusterings in a space that encompasses a continuum of states for some of which we have no commonly accepted labels. (Currie and Jones 2006, p. 312)

Although Currie and colleagues stop short of developing a positive account from this reasoning, the passage hints at a more revisionary form of non-doxasticism which depends on rejecting the ability of the categories of folk psychology to properly characterize delusional states. This idea, in turn, can result in either attributing to the delusional subject a *hybrid* state somewhere between belief and imagination, or in a more nuanced view of belief which postulates that not all cases of attribution will yield a yes or no answer to the question 'Does the subject *believe* that *p*?'.

3.2.2 Bimagination

In 'Imagination, Delusion and Self-Deception', Andy Egan (2009) proposes that delusions are instances of a novel attitude somehow intermediate between imagination and belief, which he calls *bimagination*.[10] What this means is that this hybrid attitude would possess some of the distinctive features of believing, and some of the distinctive features of imagining. As Egan observes, 'Delusions are not happily classified as either straightforward cases of belief or straightforward cases of imagining' (2009, p. 276). If on the one hand, classifying delusions as paradigmatic cases of belief is problematic because it predicts that delusions ought not to display the sorts of circumscription and evidence-independence that they apparently display, on the other hand, classifying them as paradigmatic cases of imagination is problematic because it predicts that they should display *more* circumscription and evidence-independence than they apparently display.

> What would be nice would be to be able to say that the attitude is something in between paradigmatic belief and paradigmatic imagination—that delusional subjects are in states that play a role in their cognitive economies that is in some respects like

[10] Egan (2009, p. 275ff.) also approaches the equally difficult task of characterizing self-deception, proposing that it should be understood as an attitude intermediate between belief and desire: 'besire'.

> that of a standard-issue, stereotypical belief that p, and in other respects like that of a standard-issue, stereotypical imagining that p. (2009, p. 268)

Is such a mongrel, neither-fish-nor-fowl kind of propositional attitude feasible? Egan's argument for making room in our cognitive theories for hybrid attitudes, against possible opponents who might object on principle to the promiscuous proliferation of mental attitude types, is derived from the fact that (at least some) functional roles performed by beliefs as well as by imaginings are not a package deal. Thus he argues that it is a mistake to think you cannot have the origin of an imagining and the behavior-guiding role of a belief, or a belief-like behavior guiding role here and an imagination-like behavior guiding role there, or a belief-like origin and an imagination-like updating policy, etc. To illustrate this, he invites us to consider the sort of case of inconsistent belief that David Lewis discusses in 'Logic for Equivocators'.

> I used to think that Nassau Street ran roughly east-west; that the railroad nearby ran roughly north-south; and that the two were roughly parallel. ... So each sentence in an inconsistent triple was true according to my beliefs, but not everything was true according to my beliefs. Now, what about the blatantly inconsistent conjunction of the three sentences? I say that it was not true according to my beliefs. My system of beliefs was broken into (overlapping) fragments. Different fragments came into action in different situations, and the whole system of beliefs never manifested itself all at once. The first and second sentences in the inconsistent triple belonged to-were true according to-different fragments; the third belonged to both. The inconsistent conjunction of all three did not belong to, was in no way implied by, and was not true according to, any one fragment. That is why it was not true according to my system of beliefs taken as a whole. Once the fragmentation was healed, straightway my beliefs changed: now I think that Nassau Street and the railroad both run roughly northeast-southwest. (1982, p. 436)

But even if we accept (as I think we should) a less restrictive, "boxological" view of mental attitudes, does this warrant the kind of attribution Egan has in mind? In other words, does the fragmentation of functional roles justify the use of labels such as 'bimagination'? It is easy to get the impression that Egan wants to have his cake and eat it, too. One problem I see with Egan's account is that bimagination seems so *ad hoc* as to elicit the following question: if the promiscuous proliferation of propositional attitude types is

proportional to the variety of possible functional roles, then why not just say that *delusion itself* is a type of propositional attitude with the characteristics that it has? And why stop there: wouldn't every subtype of delusion, such as Cotard or Capgras, be ultimately characterized as its own kind of propositional attitude? How is this type of characterization *informative*?

The fact is that, in the end, Egan's approach doesn't do justice to the intuition expressed by Currie and Jones. Rather than developing the idea that delusions are best characterized as 'vague clusterings in a space that encompasses a continuum of states for some of which we have no commonly accepted labels' (Currie and Jones 2006, p. 312), Egan ends up trying to fit delusion into a categorically distinct class that he suggests we should label 'bimagination,' whereas the legitimate conclusion from the premise of functional role fragmentation is a more nuanced view of the attitudes wherein the possession of contradictory dispositions can be made to make sense. As we will see in section §5, such an account is available to us.

3.3 Problems for both sides of the debate

The discussion of the two alternative, non-doxastic characterizations above highlights at least two important problems which may be endemic to the whole debate about whether or not delusions are beliefs. The first problem concerns the legitimacy of the debate itself. The second concerns the legitimacy of the claims being made by both sides.

3.3.1 A terminological dispute?

Does the debate between doxasticists and non-doxasticists turn on facts about the human mind? Authors such as Tim Bayne (2010) and Richard Dub (2013) have recently noted that it is easy to get the impression that there is nothing *substantive* being achieved by the positive proposals we have examined so far and, furthermore, that the question 'Are delusions beliefs?' might only appear to be answerable on the surface. Whether the dispute between the two sides is merely terminological is a point that deserves attention insofar as it pertains to the possibility of actually answering the question 'Are delusions beliefs?'. If the question cannot be properly answered this would have important consequences for how we should go about building a scientific theory of delusion. Thus, Bayne observes:

> Both parties agree that the functional role played by anomalous states (such as delusions) differs from that of paradigm (ordinary) beliefs. Those who are sympathetic to the doxastic account,

> such as [Marga] Reimer (2010), add that this difference is not so marked as to exclude delusions from the doxastic realm altogether, whereas those who reject the doxastic account, such as [Andy] Egan (2009), hold that although the functional role of delusions may be belief-like, it is not sufficiently belief-like for delusions to qualify as beliefs. But without an account of the functional role of belief it is not clear whether this is really a debate about how best to understand delusions, as opposed to a debate about how to use the term 'belief'. (2010, p. 332)

Similarly, Dub (2013, p. 82) points out that the question might only be settled by deciding how to use the *words* 'delusion' and 'belief' rather than about what delusions and beliefs *are*. Toward that end, it would be useful to have some sort of diagnostic test by which we could find out whether or not a debate turns on mere terminology. David Chalmers (2009, p. 88) offers the following: check whether the dispute disappears after two different senses of the problematic term are distinguished. At first sight, this test suggests that 'Are delusions beliefs?', like 'Is a cucumber a fruit?' and unlike 'Is gold a metal?' is merely terminological. As I will go over in more detail below, doxasticists seem to apply the word 'belief' to a set of psychological states that includes delusions, whereas non-doxasticists restrict the application of the word to a smaller set of psychological states which excludes delusions,[11] just as botanists apply the word 'fruit' to a wider range of objects than are understood to be fruits in a culinary sense.

However, Chalmers's test may be rightly seen as too unsophisticated. It invites us to distinguish different senses of a term, but it is not that easy to know when different senses are in play and there exists the possibility that conversationalists may converge on the same meanings even when using their words differently. Thus, Dub provides an heuristic that does not involve having to scrutinize the meanings of the terms being used: a dispute is merely terminological if and only if it is not possible to have the dispute without using whatever words are apparently troublesome.

> In reformulating the debate without using the taboo word, one will have to resort to the redescriptions, paraphrases, and translations, but we do not take on any contentious stance about whether these redescriptions are legitimate "senses" of the now-taboo word

[11] Remember the passage from Currie and Ravenscroft: 'If someone says that he has discovered a kind of belief that is peculiar in that there is no obligation to resolve or even to be concerned about inconsistencies between these beliefs and beliefs of any other kind, then the correct response to him is to say that he is talking about something other than belief' (2002, p. 176).

> or are related to its semantic content in any particular way. (2013, p. 84–85)

Though by no means foolproof and far from *defining* what a terminological dispute constitutes, Dub's diagnostic test does suggest that doxasticists and non-doxasticists are talking past one another. With this diagnostic test in hand, we can examine the two potentially problematic terms in the assertion 'delusions are beliefs', namely 'delusions' and 'beliefs'. In the following subsection, I will focus on potential indeterminacies in the word 'delusion' that give rise to an independent charge to both doxasticism and non-doxasticism, namely, that of making unjustified general claims. In the next section, I will will focus on potential indeterminacies in the word 'belief' that put in jeopardy the project of pigeonholing delusions as being definitely beliefs or definitely not beliefs.

3.3.2 Overly general claims

The characteristic circumscription that functions as the second premise in the argument from functional role against doxasticism, though certainly observed in many cases of delusion, is not a feature of *all* delusions. Just as there are examples of the failure of delusions to be integrated with the subject's beliefs, actions, and emotions, there are also cases that *do* display such integration and, therefore, lend support to the attribution of belief (Bayne and Pacherie 2005). This can easily be established with data from both empirical studies and first-person accounts of delusion.

With regard to empirical data, for example, a review of 260 cases of delusional misidentification by Hans Förstl and colleagues (1991) found that physical violence had been noted in 18% of cases. Andrew Young and Kate Leafhead (1996) note that all their Cotard patients displayed at least some measure of congruent behaviors, such as refusing to move, to eat, or to shower. J.M. O'Dwyer (1990) reports that erotomania patients commonly act on the basis of their delusion. And Simon Wessely and colleagues (1993) note that 77% of a total of 59 delusional patients acted on their delusions in the month prior to admission. Therefore, circumscription objections have only the power to undermine the *generality* of a doxastic characterization of delusions, without thereby establishing the generality of an non-doxastic characterization—especially because the empirical evidence just mentioned fits the doxastic model better, thus undermining the possibility that either Currie's or Egan's account could work as a general characterization of delusion. So if doxasticism cannot provide a general account because it fails to include the cases to which non-doxasticists point to, the reverse is also true and, thus, no positive

morals can be extracted from the debate. But here is an important *negative* moral: the heterogeneity of the class of delusions puts pressure on the very possibility of anyone ever arriving at a characterization that is at once general and precise.

Still with respect to delusional states that are *not* circumscribed as non-doxasticists paint delusion to be, consider the following testimony by Esmé Weijun Wang, a writer responsible for what is perhaps the only extant account of the experience of Cotard delusion (quoted with permission).[12]

> In the beginning of my own experience with Cotard's delusion, I woke my husband before sunup. Daphne, our dog, stirred, began thumping her papillon-mutt tail against the bedsheets. I'd been in my studio, but now I was shaking my husband, and I was crying with joy.
>
> 'I'm dead,' I said, 'and you're dead, and Daphne is dead, but now I get to do it over. Don't you see? I have a second chance. I can do better now.'
>
> Chris said, gently, 'I think you're alive.'
>
> But this statement, of course, meant nothing. It was his opinion, and I had my solid belief. I can state that the sky is green, but will you see it as such? I felt buoyant at the belief that I was getting a second chance in some kind of afterlife—it caused me to be kinder, to be more generous. I wasn't irritated by problems with computer downloads. I was sweet to telemarketers. It was true that I was dead, but I believed it made sense to play-act normalcy, or rather, an improved version of normalcy, because of the additional belief that I was in an afterlife. According to the logic of my delusion, this afterlife was given to me because I hadn't done enough to show compassion in my "real" life; and though I was now dead, my death was also an optimistic opportunity. (Wang 2014)

[12] First-person accounts of monothematic delusions are very uncommon. Indeed, the only book-length first-person account of monothematic delusion that I know of is the splendid *A Leg to Stand On*, in which the neurologist and writer Oliver Sacks describes his recovery after a fall in a remote region of Norway in which he injured his leg. Following surgery to reattach his quadriceps muscle, Sacks experienced a period in which his leg no longer felt a part of his body. He describes his confusion, seeing the 'disowned' plastered limb: '[the leg] became a foreign, inconceivable thing, which I looked at, and touched, without any sense whatever of recognition or relation. It was only then that I gazed at it, and felt I don't know you, you're not part of me, and, further, I don't know this "thing," it's not part of anything. *I had lost my leg*' (Sacks 1984, p. 53, my emphasis).

Note that Wang's conviction that she was dead was *not* inferentially circumscribed (or at least not completely), since she also formed the coherent conviction that she was experiencing an afterlife—likely an abductive explanation of the unshakeable conviction (or rather, *fact*) that, although dead, she remained a subject of experiences. Moreover, her delusional convictions had behavioral and affective consequences, leading her to verbally affirm that she was dead, to be unencumbered by petty problems, and to rejoice at the second chance she had been given. While members of one of the sides of this debate can (and do) summon examples of first-person accounts of schizophrenia,[13] for example, to illustrate the point that at least some delusions are not belief-like, the upshot of the considerations above is that delusions are highly heterogeneous and, thus, it should come as no surprise that some delusions are more belief-like, while others depart from stereotypical beliefs. What we need, it seems, is an account of delusion that embraces this heterogeneity and strives for precision without losing sight of the fact that we are dealing with a class of phenomena that might very well not be amenable to sweeping general claims.

3.4 The limits of folk psychology

Responding to the question of whether non-linguistic animals have beliefs, Stephen Stich once paraphrased his young son in saying 'A little bit they do. And a little bit they don't' (1979, p. 28). From what has been discussed so far, the response to the question of whether delusions are beliefs should fall along the same lines: 'a little bit they are, a little but they are not' (Bayne 2010). However, rather than trying to create new labels to fit borderline phenomena, as Egan does, we should pursue a nuanced account that at once recognizes the limits inherent in folk-psychological categories and provides us with a way to talk intelligibly and responsibly about phenomena which can't be made to fit such categories.

3.4.1 In-between believing

H.H. Price, in his famous series of lectures on belief, discussed the not uncommon phenomenon wherein a person may systematically feel himself to be and act as if he were fully committed to p in one set of circumstances, while systematically feeling and acting as if the opposite were true in others. He called this 'half-belief' (1960/1969, pp. 302-14). More recently, Schwitzgebel (2001)

[13] See, for example, the first-person account of a patient quoted by Sass (2004) in section §3.1 of chapter 1 of this dissertation.

recognized that there are countless cases in which a simple yes or no answer to the question 'Does S believe that p?' doesn't seem to be available, and that they can have a wide variety of causes. From these cases, Schwitzgebel draws the conclusion that

> For any proposition p, it may sometimes occur that a person is not quite accurately describable as believing that p, nor quite accurately describable as failing to believe that p. Such a person, I will say, is in an 'in-between state of belief' (2001, p. 76).

By way of illustration, he offers three examples stemming from three different causes, which are neither meant nor thought to be exhaustive. The first is *gradual forgetting.* It concerns the ubiquitous case in which someone forgets, say, an old colleague's last name. Years ago, you knew your colleague's full name. Now, you can only remember his first name (and, perhaps, the first letter of his last name). Years from now, you probably won't remember his name at all. So the belief that your colleague's name was Konstantin Guericke was fully present when you were in college, and will be fully absent when you are eighty years old. The question then is, what is the state you're in right now? Schwitzgebel asks: 'is it plausible to think that in the years between there was a discrete moment before which I absolutely had this belief and after which I absolutely did not? At some point during the course of forgetting, I must be *between believing and failing to believe* that his last name is Guericke (or whatever)' (2001, p. 77, my emphasis). Arguably, we spend most of our lives in such an in-between state.

His second example is derived from our *failure to think things through.* Think of a school teacher who mentions prime numbers in her lessons, correctly listing the lower primes 2, 3, 5, 7, 11 etc. Now, when she is asked about or decides to offer the definition of 'prime number', she typically says that a prime number is any positive integer that can be divided evenly only by 1 and itself. This definition is not correct, however, since the number 1 is a positive integer evenly divisible only by 1 and itself, but it is not a prime number. On the other hand, if you asked the school teacher if 1 is a prime she would promptly answer that it isn't. So now the question is, does she believe that all positive integers which are evenly divisible only by themselves and 1 are prime? We have reasons to answer in the affirmative, for instance, she would never list 1 as a prime number. But we also have reasons to answer in the negative, for instance, the occasions on which she would be disposed to offer a correct definition of primes are few. For this reason, Schwitzgebel claims 'the most careful and accurate description of her would neither simply ascribe the belief to her nor simply deny it of her' (2001, p. 77).

Finally, there is *variability with context and mood.* Here, Schwitzgebel evokes a familiar example in the same vein as Price's famous case of the half-believing theist. Price suggests the case of someone who on Sundays bears all the subjective and objective marks of someone who believes that there is a God, but who on weekdays bears none of them. Schwitzgebel, on the other hand, suggests the case of someone who, in certain moods and in certain contexts, bears all the subjective and objective marks, and who, in other moods and contexts, doesn't. (The latter spectrum may include circumstances from those of weakened confidence, as when someone thinks of God as 'a beautiful metaphor', to those where confidence is removed completely from recognition or memory.) Though he may be a regular Sunday churchgoer, he does not feel the urge to defend himself or his religion when, for example, his atheistic friends mock religious belief. In fact, at such moments (especially on weekdays), he may even find himself mildly convinced of the incongruousness of theistic dogma. How can we decide, then, whether he believes that God exists? Once again, Schwitzgebel makes the point that a simple yes or no answer would be misleading.

> One might say that his beliefs change from occasion to occasion—that as he is grousing about the church social, he does not believe that God exists; as he is rejoicing in the magnificence of spring, he does believe—but most of the time he is doing neither: he is eating breakfast or mowing or writing code and not giving the matter any thought. At such moments he may be simultaneously disposed to marvel at the wonder of creation if a robin were to fly past and to embrace atheism if Madge were unexpectedly to drop by. (2001, p. 78)

The widespread presence of problematic circumstances for belief-ascription such as these encourages an account of belief that allows us to talk intelligibly about such in-between states—an account that allows us to say more than just that the subject 'sort of' believes something. Given the notion that there is a continuum ranging from complete absence to complete presence of any given belief, a probabilistic treatment might be thought to manage cases of in-between believing. According to such an account, a person's beliefs would be characterized by a degree of confidence ranging from 0 (i.e. absolute confidence in the falsity of p) to 1 (i.e. absolute confidence in the truth of p), with 0.5 in between—perhaps representing suspension of judgment or a state of

skeptical doubt.[14] Such an approach may be thought to account for at least some of the cases because we could assign our half-believing theist, for example, with a degree of confidence of 0.7 or 0.8. However, this would consist in a gross oversimplification of the kind of uncertainty or wavering present in the cases discussed. The school teacher and the half-believing theist cannot be properly described as simply fluctuating between different degrees of confidence, since they are, 'at a single time, disposed quite confidently to assert one thing in one sort of situation and to assert its opposite in another' (Schwitzgebel 2001, p. 79). Nor can the process of gradually forgetting someone's last name be properly translated into a slow decline in one's confidence in the truth of some proposition. A purely probabilistic approach fails to capture the vast array of detail present in these cases.

Furthermore, it would seem that traditional representational ("boxological") accounts of belief cannot provide a way of successfully dealing with in-between belief states either. Indeed, to suggest that someone is in an in-between representational state appears even more unnatural than the probabilistic strategy would have it. Most talk of belief as representation makes out belief to be a *categorical* state—having a belief that p is something like having the sentence p inscribed in one's 'belief box' in the language of thought, according to one popular account. The metaphor must be pushed, though, if representationalists wish to embrace the very plausible presence of halfway states. Schwitzgebel points out that for that, however, they risk making a caricature of their own account by incorporating, say, explanations of gradual forgetting in terms of a sentence slowly 'losing its color', etc. To avoid the far-fetched claim that sentences either are or aren't inscribed in the belief box, then, Schwitzgebel claims that representationalists are left with the burden of coming up with helpful ways of describing in-between cases in representational terms.

Schwitzgebel opts for pursuing a more flexible explanation of the nature of belief and belief-ascription by appeal to a revision of Gilbert Ryle's dispositionalism. Ryle argued that to believe something is simply to be disposed to do and feel certain things in appropriate situations. To use his own example, to believe that the ice you're skating on is dangerously thin is, in his words,

> to be unhesitant in telling oneself and others that it is thin, in acquiescing in other people's assertions to that effect, in object-

[14] 'Probabilistic' is the way I have chosen to put it. Schwitzgebel (2001, 2002) chooses the word 'Bayesian' but there is nothing specifically bayesian about the view he describes. The notion that belief comes in different degrees of confidence is part of every probabilistic account of belief (Ramsey–DeFinetti's, for instance). However, since there is no mention of *conditional* probabilities (conditional, that is, upon prior beliefs), nothing warrants Schwitzgebel's choice of label.

> ing to statements to the contrary, in drawing consequences from the original proposition, and so forth. But it is also to be prone to skate warily, to shudder, to dwell in imagination on possible disasters and to warn other skaters. It is a propensity not only to make certain theoretical moves but also to make certain executive and imaginative moves as well as to have certain feelings. (1949, pp. 134-5)

A person who has the dispositions described in Ryle's example matches what Schwitzgebel calls a *dispositional stereotype*. By a stereotype, he means a cluster of properties we are apt to associate with something—be it an object, a class, or a property. An example he adapts from Hilary Putnam (1975) is that of the stereotype of a tiger, whose properties include being striped and having four legs, among others. This doesn't mean, of course, that a three-legged tiger without stripes is not a tiger. It only means that such a tiger wouldn't be a stereotypical one. Furthermore, the accuracy of stereotypes varies greatly in degree, so that the more or less objects instantiate their stereotypical properties, the more or less accurate the stereotype will be.

A dispositional stereotype is simply a stereotype whose elements are dispositional properties.[15] Many familiar stereotypes are dispositional, such as personality traits. For example, being impulsive is (something like) being disposed to act without thinking things through; being sympathetic is (something like) being disposed to easily putting oneself in someone else's position; etc. Just like having a personality trait is matching a stereotype, Schwitzgebel claims, so too is having a belief. As a consequence, the list of dispositions associated with a given belief is as indefinite as that of having a particular personality trait, and won't be linked to it explicitly by a conscious effort. The most fruitful way of thinking about dispositional stereotypes is, rather, as consisting of clusters of dispositional properties (which we associate with particular stereotypes). We associate specific clusters of behavioral, cogni-

[15] Schwitzgebel characterizes dispositions by means of conditional statements of the form 'If condition C holds, then object O will (or is likely to) enter (or remain in) state S' (2002, p. 250). The latter sentence states explicitly that there is a law-like connection between being in condition C and entering (or remaining) in state S, something which no strictly indicative conditional would have the force to express ('If P then Q' iff 'Not-P and/or Q': no law to be found there). O's entering S is the *manifestation* of a disposition, whereas C is the *condition of manifestation*, and the event of C's obtaining is the *trigger*. Therefore, O will have the relevant disposition if and only if the corresponding conditional statement is true. Thus we may speak of dogs having the disposition to wag their tails when excited because when they are excited (the trigger), they wag their tails (the manifestation)—which, please note, does not mean that every dog is such that it is not excited and/or its tail wags—which would be true of a dog which is excited without wagging its tail—all indicative conditionals whose antecedent is true being alike trivially true.

tive, and phenomenal dispositions with given beliefs and expect them to be manifested in standard situations. We thereby attribute a belief to the subject if he conforms to the associated stereotype in standard situations and if his deviations from the stereotype are readily explainable or excusable by appeal to some non-standard feature of the situation in which they occur.[16]

3.4.2 The sliding scale approach to delusion

In the most difficult cases for ascription—amongst which delusion certainly has a special place—the communicative demands on the attributor may not successfully determine whether or not it is appropriate to describe the subject as believing the content of what they profess to believe. Schwitzgebel (2012) argues that cases like these, in which the set of ascribable dispositions available to the interpreter is such a "mixed bag," leave us only with the option of *specification*—that is, describing how the subject's dispositions conform to the stereotype for the belief in question and how they deviate from it. There will be times, then, when *withholding* the use of ascriptive language is going to be preferable so as not to mislead one's audience. Such cases are those in which the observable deviations raise questions regarding both the content of the subject's attitude, and the nature of the attitude itself.

So if there is no way to decide whether something is determinately a case of belief, our move should be to allow *some* indeterminacy in our belief talk, for fear that we should abandon it altogether. In keeping with this, Schwitzgebel offers a mostly neglected way for handling delusional states (or at least those which defy ascriptive language and practice). He suggests that 'believes that p' should be treated as a vague predicate admitting of vague cases:

> In in-between cases of canonically vague predicates like 'tall', the appropriateness of ascribing the predicate varies contextually, and often the best approach is to refuse to either simply ascribe or simply deny the predicate but rather to specify more detail (e.g., 'well, he's five foot eleven inches'); so too, I would argue, in in-between cases of belief. (2012, p. 15)

Rather than supporting the view that delusions are beliefs (or at least that some of them are), however, all that Schwitzgebel's view can really offer is a pragmatic license to talk about delusions as beliefs whenever this is not apt to mislead our intended audience, and whenever there is no better alternative. Therefore, Schwitzgebel's view is not fully a doxasticist view

[16] Bayne and Pacherie (2005, p. 185) cite a fear of involuntary commitment, for example, to account for the failure of some patients to act on their alleged beliefs.

about delusion. Besides, it is conceivable that among the many cases that defy belief-ascriptive language there might be some cases of delusion that imagining-ascriptive language is better suited to describe (even if in localized instances, for the benefit of particular audiences). The fact that belief-ascriptive shorthand caters to the context and interests of the attributors defeats the doxasticist's purpose of defending a full-blooded doxastic view of delusions by appeal to dispositionalism about belief, as has been proposed (Bayne and Pacherie 2005).

But where does vagueness get us? Bortolotti (2010, pp. 20–1) dismisses this kind of 'sliding scale' approach on the grounds that, by not giving a straightforward answer to the question 'Does the patient believe that p?', it is unable to characterize precisely whether the patient's actions are intentional, which complicates issues of ethical and policy-guiding import.[17] However, Schwitzgebel retorts that this is not nearly enough reason to discard the approach without more ado, since its proponents might just as well suggest that 'in many cases of delusion it *shouldn't* be straightforward to assess intentionality, and that the ethical and policy applications *are* complicated, so that a philosophical approach that renders these matters straightforward is misleadingly simplistic' (2012, p. 15). Ironically, toward the end of her book, Bortolotti hints at the in-between approach we have been discussing when she writes:

> Rarely do we have these clear-cut cases ... Most of the delusions we read about, and we come across, are integrated in the subject's narrative, to some extent, and with limitations. They may be excessively compartmentalized, for instance, or justified tentatively. That is what makes it so difficult to discuss the relationship between delusions, subjects' commitment to the content of the delusion, and autonomy. As authorship comes in degrees, so does the capacity to manifest the endorsement of the delusional thought in autonomous thought and action. (2010, p. 252)

As Schwitzgebel observes, from the fact that Bortolotti (2010, p. 242) regards authorship and endorsement as necessary for belief, it seems to follow that in the quoted passage she is acknowledging that many actual delusions are

[17] Another practical reason that Bayne and Pacherie (2005) consider is the fact that effective therapeutic treatments are formulated in terms of patients's beliefs. For example, cognitive behavioral therapy (CBT), an important form of therapy for delusions (Dickerson 2000), involves questioning the consistency and plausibility of the patient's delusions (Chadwick et al. 1996). This form of therapy, the authors argue, is consistent with doxasticism inasmuch as the therapist treats the delusional patient as a *believer* of p, proceeding to gently invite the patient to question whether p is ought to be believed.

in-between cases of belief. This wavering on Bortolotti's part is symptomatic of an increasingly widespread (Hamilton 2007), if latent, perception of which a recent formulation can be found in the words of Tim Bayne: 'there may not be enough determinacy in our ordinary conception of belief for there to be a fact of the matter as to whether many belief-like states are really beliefs or not' (2010, p. 332).

Thus, Schwitzgebel concludes that 'when a person deviates too much from the causal-functional patterns in behavior and cognition characteristic of belief, the assumptions inherent in the practice of belief ascription start to break down; and then we have to either abandon belief talk or allow for some indeterminacy in it' (2012, p. 15). As we have just seen, Schwitzgebel opts for allowing indeterminacy in belief talk, and I agree that that is convenient enough for everyday purposes where precision is not a definitive issue. But what about when we are attempting to arrive at a scientific theory of the relevant phenomena? How does allowing for indeterminacy in belief talk help us achieve a characterization of delusion, let alone an explanatory theory of it? Though Schwitzgebel offers an important moral, his characterization, as all other characterizations we have discussed, does not offer anything by way of explanatory power. In the next and final section, I will suggest that the underlying problem with all characterizations discussed so far lies with their insistence on their single-minded focus on person-level psychology and their inability to integrate with other relevant levels of explanation. We should pursue an explanatory theory of delusion that does away with terminological disputes once and for all and which privileges whatever levels of description wherein precision can actually be achieved.

3.5 Moving past the debate

The characterizations of delusion assessed so far are found lacking in two further respects, which, I argue, deal a fatal blow not to any particular characterization, but to the project of explaining delusion by investing in folk-psychological terminology. First, by focusing too hard on which propositional attitude delusional subjects are supposed to hold with respect to the content of their delusions, they fail to make any progress in addressing the question of how the delusional patient experiences his or her delusions. Second, they fail the main conceptual challenge in offering a characterization of delusion, namely, to provide a unifying framework that would make it easier to look downwards to the neural mechanisms underlying delusions, thus failing to carry explanatory weight.

The theories discussed so far, all of which reduce delusion to a single

propositional attitude (however boxological or nuanced), face the charge of being descriptively inaccurate when attention is given to the experience of delusional subjects. As we have seen in §3.2, some first-person accounts, such as Esmé Weijun Wang's account of her experience of Cotard's, may function as evidence against non-doxasticism inasmuch as the circumscription invoked by non-doxasticists to attack the doxastic status of delusion is absent in at least some cases. This should not, however, be immediately seen as a victory for the doxastic side, insofar as doxasticism faces a similar problem with respect to a variety of cases. First-person accounts of schizophrenia in particular suggest that the question of how the delusional patient takes the world to be will hardly be answerable by referring to a determinate belief (or other kind of attitude) with respect to a proposition. Indeed, the more complex and florid the delusion or delusional system of the subject, the clearer this point seems to become. Consider the celebrated case of Daniel Paul Schreber, whose *Memoirs of My Nervous Illness* (Schreber 1903) inspired Jaspers's theory of the incomprehensibility of delusion and which has been the focus of an extensive case study by Louis Sass (1994).

> I can put this point briefly: everything that happens is in reference to me. ... Since God entered into nerve contact with me exclusively, I became in a way for God the only human being around whom everything turns, to whom everything that happens must be related and who therefore, from his own point of view, must also relate all things to himself. (Schreber *apud* Sass 1994, p. 61)
>
> This completely absurd conception, which was at first naturally incomprehensible to me but which I was forced to acknowledge as a fact through years of experience, becomes apparent at every opportunity and occasion. For instance when I read a book or newspaper one thinks that the ideas are my own; when I play a song or opera arrangement for the piano, one thinks that the text of the song or opera expresses my own feelings. (ibid.)
>
> I have to add that the female characteristics which are developing on my body show a certain periodicity at increasingly shorter interval. The reason is that everything feminine attracts God's nerves. Hence, as often as one attempts to make the female characteristics which are evident on my body recede by miracle; the effect is that the structures which I call "nerves of voluptuousness" are pushed a little under the surface, that is to say are not distinctly palpable on the skin, my bosom becomes a little flatter, etc. But after a short time the rays have to approach

> again, the "nerves of voluptuousness" (to retain this term) become more marked, my bosom bulges again, etc. Such changes occur at present in as short a period as a few minutes. (ibid., 123)

Rather than saying that the delusional subject believes, imagines, "bimagines," possesses some of the stereotypical dispositions of belief but not others, etc., one might as well say that the individual experiences the delusion 'as a subjective reality or framing condition for the living of life as the person whom they are' (Mullen and Gillett 2014, p. 35) in that nothing is to be gained from such a characterization in terms of actually *explaining* the condition. Indeed, such testimonies do not give rise to the question 'Did Schreber *believe* such and such?' so much as to the etiological and explanatory questions 'What gave rise to Schreber's experiences?' and 'Why did he interpret them the way he did?'. As Tim Bayne observes, even if the concept of belief were sufficiently precise, 'it is a further question as to *why we should care* about whether delusions are anomalous beliefs or some in-between state such as bimaginations. Arguably, what matters for many purposes is the question of what functional role delusions actually play, rather than whether this functional role falls within the boundary of belief or not' (2010, p. 332). So, in addition to providing overly general characterizations that are not up to the task of precisely describing the delusional subject's attitude toward their delusions, it is worth asking ourselves if and why the language of folk psychology is apt to play a relevant role in an *explanation* of delusion (and, for that matter, other cognitive phenomena). The vocabulary of folk psychology, though a useful tool for conceptualizing and dealing with ourselves and others, abstracts entirely from cognitive and neural processes, thereby putting in jeopardy the possibility of an integrative explanation of the phenomena.

Jakob Hohwy (2013, p. 57) notes that the explanatory challenge involved in devising a characterization of delusion is to provide a unifying framework that would make it easier to look downwards to the cognitive and neural mechanisms underlying delusions. Characterizations that invest in folk-psychological terminology, thus, being abstractions from lower-level processes, fail to provide us with such a unifying framework and hinder a multilevel explanation of delusion. For this reason, Philip Gerrans (2014) suggests we take the advice of Dominic Murphy and let the cognitive neuroscience determine our characterization of psychiatric disorder in general and delusion in particular: 'we arrive at a comprehensive set of facts about how the mind works, and then ask which of its products and breakdowns matter for our various projects' (Murphy 2006, p. 105). However, this should not be misunderstood as entailing that the appropriate level of explanation is

the *lowest*-level, i.e., molecular biology. To the contrary, Murphy advocates explanatory pluralism to the effect that there is no fundamental level, and explanations in cognitive neuropsychiatry must include references to factors that span all levels—from molecular biology to the cognitive and social sciences. As Gerrans puts it, 'no part of biology or psychology has proprietary rights to psychiatric explanation' (2009, p. 113). The suggestion that we should take our lead from cognitive neuroscience and not person-level folk-psychology is then perfectly at home with such an explanatory pluralism and is only meant to drive home the point that there is no place for such abstractions in a causal, mechanistic explanation of delusion (though there might perfectly well be a place for 'belief' and the like in other pragmatic contexts, such as therapeutic and forensic).

Conclusion

In the preceding sections, I have attempted to elucidate that both doxasticism and non-doxasticism fail to characterize the functional role of delusions while at the same time being unable to play a role in the explanation of these phenomena. Both sides of the debate offer characterizations that are easily seen to downplay the immense variety in said functional role, and the debate ultimately turns on how its members apply the words 'delusion,' 'belief,' etc., thus consisting of a merely terminological dispute. Though a more nuanced view of belief wherein mental states are more or less belief-like instills a healthy skepticism towards the precision of folk-psychological concepts, I have argued that it fails to be of any use in building a theory of delusion that may be able to bridge different levels of explanation, such as the phenomenology and neurobiology of delusion. Thus, I advocate moving past the question 'Are delusions beliefs?' and their description as propositional attitudes toward the description of the processes that generate delusion, with a view toward explaining, rather than explaining away, the personal-level aspects of the phenomenon that have been made inscrutable by doxastic terminology.

Conclusão

By working with scientists I get a rich diet of fascinating facts to think about, but by staying a philosopher I get to think about all the theories and experiments and never do the dishes.
Daniel Dennett

Ao longo desta tese, espero ter-me desincumbido de pelo menos três modestas tarefas. A primeira foi apresentar ao leitor uma família de fenômenos mentais—a saber, os delírios estudados pela psiquiatria—e algumas das principais dificuldades que têm sido enfrentadas nos esforços de compreensão e explicação destes. A segunda foi apontar falhas nas principais tentativas de resposta a algumas dessas dificuldades—a saber, o estatuto de espécie natural e o estatuto de crença do delírio—especialmente no que diz respeito à adequação dessas respostas ao desenvolvimento de uma teoria científica do delírio. A terceira foi esboçar hipóteses de trabalho que visam reparar as falhas das soluções discutidas previamente e oferecer suporte teórico à explicação dos fenômenos investigados.

A primeira hipótese diz respeito ao estatuto de espécie natural do delírio. Ostensivamente, a investigação sobre os critérios para a respeitabilidade e relevância científica de categorias torna claro que o essencialismo é rigoroso demais para fazer sentido das nossas práticas científicas, nas quais não há qualquer dúvida de que espécies biológicas, por exemplo, constituem categorias que dão suporte a generalizações indutivas e explicações mecanicistas. Todavia, categorias da psiquiatria não possuem imediatamente o mesmo estatuto de categorias da biologia. Quando doenças mentais e sintomas psicopatológicos constituem grupos estáveis de propriedades, estas podem ser consideradas, minimamente, espécies práticas. Se, por outro lado, conseguirmos oferecer explicações da covariação de alguns desses grupos de propriedades em termos de mecanismos causais comuns, então essas categorias psiquiátricas seriam elevadas ao mesmo estatuto concedido a categorias da biologia. Todavia, teorias cognitivas da deteção e atribuição de doenças mentais apontam para uma importante dependência dessas categorias com respeito a juízos

intuitivos sobre o que é normal e isso, por sua vez, coloca o ônus da prova sobre aquele que pretende reivindicar o estatuto de espécie natural do delírio.

Porém, mesmo que a constatação de que a categoria do delírio se trata de uma formalização de intuições do senso comum não remova *imediatamente* a possibilidade de que esta categoria seja, ao fim e ao cabo, uma espécie natural, o projeto de reivindicação da categoria do delírio se mostra metodologicamente questionável. Desse modo, ofereci uma via intermediária entre relegar todos os delírios a meras espécies práticas e insistir no estatuto de espécie natural da categoria genética do delírio: juntamente com a categoria genérica do delírio, alguns subtipos de delírio serão confinados ao estatuto de espécies práticas, enquanto alguns subtipos de delírio serão reivindicados como espécies naturais no sentido menos exigente segundo o qual espécies biológicas o são. Importantemente, essa hipótese e sugestão metodológica evita o embaraço causado pelo reconhecimento de que, se vários tipos diferentes de mecanismos são responsáveis pelos delírios, o ônus da prova recai sobre aquele que quiser tratar a categoria genérica do delírio ela mesma como uma espécie natural. Em outras palavras, não temos razões suficientes para presumir que uma variedade de mecanismos tornam subtipos de delírio subtipos de um mecanismo genérico, ao invés de subtipos de uma coleção heterogênea (mas mesmo assim teórica, prática e clinicamente útil) de mecanismos cujos produtos compartilham propriedades superficiais.

A segunda hipótese diz respeito ao estatuto doxástico do delírio. Ostensivamente, o exame do debate sobre o estatuto doxástico do delírio mostra que a categoria da crença, embora seja aplicada com sucesso em um grande número de situações, perde seu poder explanatório e preditivo em ao menos uma parte substancial de casos limítrofes—o que somente supreendenderia a quem erroneamente esperasse de um jargão popular um grande nível de precisão cki(uma expectativa certamente alimentada pelo amplo uso desse jargão na ciência cognitiva). Porém, não há uma linha que divida precisamente o que definitivamente é um caso de crença daquilo que definitivamente não é. Ainda, o debate sobre a natureza do delírio é complicado pelo fato de que não há consenso sobre o quanto um estado mental pode se afastar de padrões de racionalidade e integração antes que este deixe de ser denominado uma crença. Assim, delírios serão considerados racionais ou não segundo diferentes critérios de racionalidade. Portanto, 'crê que p' deve ser reconhecido como um predicado vago que admite casos vagos. Em casos intermediários de predicados vagos canônicos como 'alto,' a adequação da atribuição do predicado varia contextualmente, e muitas vezes a melhor abordagem é recusar simplesmente atribuir ou deixar de atribuir o predicado e, ao invés disso, especificá-lo em maior detalhe (e.g. 'fulano tem um metro e setenta e nove de altura'). A mesma postura pode e deve ser tomada com respeito

a casos de difícil atribuição. Nesses casos—que certamente não se limitam a casos de delírios, mas a muitos outros como vieses implícitos, negação e autoengano—as atribuições mais cuidadosas se absterão de atribuir ou negar a crença irrestritamente.

Porém, a confusão causada pela ausência de critérios precisos para a atribuição de crença e racionalidade prejudica psicólogos, clínicos e neurocientistas que laboram em busca de correlatos do delírio, como vieses cognitivos, lapsos de memória de trabalho, padrões não usuais de atividade cerebral como distribuições não usuais de receptores de dopamina ou irregularidades de processamento sináptico, entre outros. Serão estes os substratos da fixação de 'crenças' segundo princípios de inferência racionais, ou algum outro tipo de processo cognitivo? A hipótese não é, portanto, que devemos buscar uma caracterização melhor invocando *outro* tipo de atitude proposicional—que estará sujeito à mesma vagueza das outras categorias da psicologia do senso comum—mas sim que devemos buscar uma caracterização precisa dos fenômenos que queremos descrever onde esta *pode*, de fato, ser encontrada. Assim, a segunda conclusão que proponho é que não devemos deixar que nossas caracterizações do delírio (e de distúrbios psiquiátricos em geral) determinem o que ocorre nos níveis de explicação da neurociência cognitiva, mas o inverso: devemos determinar as propriedades cognitivas dos sistemas neurais envolvidos em determinado distúrbio e expressarmos a explicação em termos de processos cognitivos. Isso não acarreta um eliminativismo com respeito ao vocabulário do nível pessoal, sob condição de que este possa ser mapeado contra o pano de fundo da nossa melhor teoria cognitiva.

Um desafio conceitual pendente trata-se, portanto, de que tipo de metodologia tal mapeamento deve seguir. Uma possibilidade é que siga a metodologia intervencionista segundo a qual uma boa explicação é atingida quando podemos previsivelmente intervir e manipular componentes de um sistema. Para tanto, precisamos ter boa compreensão dos mecanismos básicos que o compreendem. Esta está ausente nas caracterizações de delírios que investem em terminologia derivada da psicologia do senso comum, pela simples razão de que uma noção como 'crença' *abstrai* de processos cognitivos e neurais sem oferecer, em troca, nada de substantivo em termos explanatórios. Assim, se concordarmos que o objetivo explanatório da psiquiatria deve ser uma explicação integrativa que nos permita passar da mera correlação à explicação causal—demonstrando como mecanismos em diferentes níveis do sujeito, do neural ao pessoal, se situam em relações de mútua manipulabilidade, por exemplo—então se torna claro que devemos abandonar caracterizações doxásticas em favor de explicações que deem conta das características dos delírios que tornam tais caracterizações incompletas e imprecisas.

Finalmente, o engajamento de filósofos com a literatura clínica sobre os

delírios e a colaboração entre filósofos e cientistas da cognição é, idealmente, uma via de duas mãos: enquanto filósofos podem se beneficiar na medida em que a literatura clínica oferece exemplos concretos com os quais teorias em filosofia da mente podem ser analisadas, estes também podem contribuir não apenas esclarecendo conceitos e tirando conclusões a partir de resultados empíricos, mas também auxiliando na construção de modelos explanatórios e sugerindo novos caminhos para pesquisa empírica. Ao longo desta tese procurei sugerir que a melhor forma em que filósofos podem contribuir para a expansão do entendimento sobre os fenômenos discutidos é resistir à tendência a formular questões e argumentos em termos exageradamente polarizantes e fiar-se nessas formulações ao explorar o domínio de investigação. Desse modo, as ferramentas de análise conceitual à nossa disposição podem servir o propósito de reparar, no seio das ciências, simplificações artificiais de fenômenos plenos de complexidades que desafiam nossas conceptualizações iniciais.

Bibliografia

Aimola Davies, A.M. & Davies, M. (2009). Explaining pathologies of belief. In M. Broome & L. Bortolotti, eds., *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford University Press.

Alexander, M.P., Stuss, D.T., & Benson, D.F. (1979). Capgras' syndrome: A reduplicative phenomenon. *Neurology* 29: 334–339.

Abed, R.T. & Fewtrell, W.D. (1990). Delusional misidentification of familiar inanimate objects. A rare variant of Capgras syndrome. *British Journal of Psychiatry* 157: 915–917.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*. Fifth edition. American Psychiatric Association.

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, & Psychology* 11(1): 1–11.

Bayne, T. & Pacherie, E. (2005). In Defence of the Doxastic Conception of Delusions. *Mind & Language* 20(2): 163–188.

Beebee, H. & Sabbarton–Leary, N. (2010). Are psychiatric kinds "real"? *European Journal of Analytic Philosophy* 6(1): 112–7.

Bell, V., Halligan, P.W. & Ellis, H.D. (2006). Diagnosing delusions: a review of inter–rater reliability. *Schizophrenia Research* 86: 76–79.

Bentall, R.P., Kaney, S., & Dewey, M.E. (1991). Persecutory delusions: An attribution theory analysis. *British Journal of Clinical Psychology* 30: 13–23.

Berke, J.H., Pierides, S., Sabbadini, A. & Schneider, S., eds. (1998). *Even Paranoids Have Enemies: New Perspectives on Paranoia and Persecution*. Routledge.

Bermúdez, J. (2001). Normativity and rationality in delusional psychiatric disorders. *Mind & Language* 16(5): 493–457.

Berrios, G.E. (1991). Delusions as 'wrong beliefs': A conceptual history. *British Journal of Psychiatry* 159: 6–13.

Berrios, G.E. (1996). *The History of Mental Symptoms*. Cambridge University Press.

Berrios, G.E. & Luque, R. (1995). Cotard's syndrome: analysis of 100 cases. *Acta Psychiatrica Scandinavica* 91(3): 185–8.

Berrios, G.E. & Kennedy, N. (2002). Erotomania: a conceptual history. *History of Psychiatry* 13: 381–400.

Bisiach, E., & G. Geminiani. (1991). Anosognosia related to hemiplegia and hemianopia. In G. P. Prigatano and D. L. Schacter, eds., *Awareness of deficit after brain injury: Clinical and theoretical issues*. Oxford University Press.

Blakemore, S.-J., Oakley, D.A. & Frith, C.D. (2003). Delusions of alien control in the normal brain. *Neuropsychologia* 41: 1058–1067.

Bleuler, E. (1911/1950). *Dementia Praecox or the Group of Schizophrenias*, trad. Joseph Zinkin. International Universities Press.

Bleuler, E. (1916/1924). *Textbook of Psychiatry*, trad. A.A. Brill. Macmillan.

Bolton, D. (2008). *What is Mental Disorder? An Essay in Philosophy, Science, and Values*. Oxford University Press.

Bortolotti, L. (2010). *Delusions and Other Irrational Beliefs*. Oxford University Press.

Bortolotti, L. (2013). Delusion. In Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*. URL = <http: //plato.stanford.edu/archives/win2013/entries/delusion/>.

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies* 61: 127–148.

Boyd, R. (1999). Homeostasis, species and higher taxa. In R. Wilson, ed., *Species: New Interdisciplinary Essays*. MIT Press.

Boyd, R. & Richerson, P.J. (1985). *Culture and the evolutionary process*. University of Chicago Press.

Boyer, P. (2001). *Religion Explained: The Evolutionary Origins of Religious Thought*. Basic Books.

Boyer, P. (2011). Intuitive expectations and the detection of mental disorder: A cognitive background to folk–psychiatries. *Philosophical Psychology* 24(1): 95–118.

Breen, N., Caine, D., Coltheart, M., Hendy, J. & Roberts, C. (2000). Towards an understanding of delusions of misidentification: Four case studies. In M. Coltheart and M. Davies, eds., *Pathologies of Belief*. Blackwell.

Brown, T.A. & Barlow, D.H. (2009). A Proposal for a Dimensional Classification System Based on the Shared Features of the DSM–IV Anxiety and Mood Disorders: Implications for Assessment and Treatment. *Psychological Assessment* 21(3): 256–271.

Browning, S. & Jones, S. (1988). Ichthyosis and delusions of lizard invasion. *Acta Psychiatrica Scandanavica* 78: 766–767.

Butler, P. (2000). Reverse Othello syndrome subsequent to traumatic brain injury. *Psychiatry: Interpersonal and Biological Processes* 63: 85–92.

Campbell, J. (1999). Schizophrenia, the Space of Reasons, and Thinking as a Motor Process. *The Monist* 82(4): 609–625.

Campbell, J. (2001). Rationality, meaning and the analysis of delusion. *Philosophy, Psychiatry, & Psychology* 8(2–3): 89–100.

Cermolacce, M., Sass, L. & Parnas, J. (2010). What is Bizarre in Bizarre Delusions? A Critical Review. *Schizophrenia Bulletin* 36(4): 667–679.

Chadwick, P.D.J., & Lowe, C.F. (1990). Measurement and modification of delusional beliefs. *Journal of Consulting and Clinical Psychology* 58: 225–232.

Chadwick, P., Brichwood M. & Trower, P. (1996). *Cognitive Therapy for Delusions, Voices and Paranoia*. Wiley.

Chalmers, D. (2009). Ontological anti-realism. In D. Chalmers, D. Manley & R. Wasserman, eds., *Metametaphysics: New Essays in the Foundations of Ontology*. Oxford University Press.

Chapman, L.J. & Chapman, J. (1988). The genesis of delusions. In T.F. Oltmanns & B.A. Maher, eds., *Delusional Beliefs*. Wiley.

Chowdhury, A.N. (1996). The definition and classification of Koro. *Culture, Medicine and Psychiatry* 20(1): 41–65.

Christodoulou, G.N. (1977). The syndrome of Capgras. *British Journal of Psychiatry* 130: 556–64.

Christodoulou, G.N. (1978). Syndrome of Subjective Doubles. *The American Journal of Psychiatry* 135(2): 249–152.

Christodoulou, G.N. (1986). *Delusional Misidentification Syndromes*. Karger.

Clark, L., Watson, D. & Reynolds, S. (1995). Diagnosis and classification of psycho pathology: Challenges to the current system and future directions. *Annual Review of Psychology* 46: 121–153.

Cleckley, H. (1941). *The Mask of Sanity: An Attempt to Clarify Some Issues About the So–Called Psychopathic Personality*. Mosby.

Coltheart, M. (2007). Cognitive neuropsychiatry and delusional belief. *The Quarterly Journal of Experimental Psychology* 60(8): 1041–1062.

Coltheart, M (2011). The mirrored–self misidentification delusion. *Neuropsychiatry* 1(6): 521–523.

Coltheart, M. (2013). On the Distinction between Monothematic and Polythematic Delusions. *Mind & Language* 28(1): 103–112.

Cooper, R. (2013). Natural Kinds. In K.W.M. Fulford, M. Davies, R.G.T. Gipps, G. Graham, J.Z. Sadler, G. Stanghellini & T. Thornton, eds., *The Oxford Handbook of Philosophy and Psychiatry*. Oxford University Press.

Craver, C. (2009). Mechanisms and natural kinds. *Philosophical Psychology* 22(5): 575–594.

Cuijpers, P., van Straten, A., Smit, F., Mihalopoulos, C. & Beekman, A. (2008). Preventing the onset of depressive disorders: a meta–analytic review of psychological interventions. *American Journal of Psychiatry* 165(10): 1272–80.

Currie, G. (2000). Imagination, delusion and hallucinations. In M. Coltheart and M. Davies, eds., *Pathologies of Belief*. Blackwell.

Currie, G. & Jones, N. (2006). McGinn on delusion and imagination. *Philosophical Books* 47(4): 306–313.

Currie, G. & Jureidini, J. (2001). Delusion, rationality, empathy. *Philosophy, Psychiatry, & Psychology* 8(2–3): 159–62.

Currie, G. & Ravenscroft, I. (2002) *Recreative Minds*. Oxford University Press.

Custance, J. (1952). *Wisdom, Madness and Folly: The Philosophy of a Lunatic*. Pellegrini & Cudahy.

Dalgalarrondo, P., Fujisawa, G., Banzato, C.E. (2002). Capgras syndrome and blindness: against the prosopagnosia hypothesis. *Canadian Journal of Psychiatry* 47(4): 387–8.

David, A.S. (1990). Insight and psychosis. *British Journal of Psychiatry* 156: 798–808.

David, A.S. (1999). On the impossibility of defining delusions. *Philosophy, Psychiatry, & Psychology* 6: 17–20.

David, A.S. & Halligan, P.W. (1996). Editorial. *Cognitive Neuropsychiatry* 1: 1–3.

Davies, M. & Coltheart, M. (2000). Introduction: Pathologies of Belief. In M. Coltheart and M. Davies, eds., *Pathologies of Belief*. Blackwell.

Davies, M., Coltheart, M., Langdon, R. & Breen, N. (2001). Monothematic delusions: Towards a two–factor account. *Philosophy, Psychiatry, & Psychology* 8: 133–58.

De Pauw, K.W. & Szulecka, T.K. (1988). Dangerous delusions: Violence and the misidentification syndromes. *British Journal of Psychiatry* 152: 91–96.

Dub, R. (2013). *Delusions, Acceptances, and Cognitive Feelings*. PhD dissertation. Rutgers University.

Dupré, J. (1981). Natural Kinds and Biological Taxa. *The Philosophical Review* 90(1): 66–90.

Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Harvard University Press.

Egan, A. (2009). Imagination, delusion, and self–deception. In T. Bayne and J. Fernández, eds., *Delusion and Self–Deception*. Psychology Press.

Ellis, H.D. & Young, A.W. (1990). Accounting for delusional misidentifications. *British Journal of Psychiatry* 157: 239–248.

Ellis, H.D., Young, A.W., Quayle, A.H. & De Pauw, K.W. (1997). Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society: Biological Sciences* B264: 1085–92.

Ereshefsky, M. (2001). *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge University Press.

Ereshefsky, M. (2009). Natural kinds in biology. In E. Craig, ed., *Routledge Encyclopedia of Philosophy*. Routledge.

Fodor, J.A. (1983). *The Modularity of Mind*. MIT Press.

Ford, J.M., Mathalon, D.H. Kalba, S., Whitfield, S., Faustman, W.O. & Roth, W.T. (2001). Cortical responsiveness during talking and listening in schizophrenia: An event–related brain potential study. *Biological Psychiatry* 50(7): 540–549.

Förstl H. & Beats, B. (1991). Charles Bonnet's description of Cotard's delusion and reduplicative paramnesia in an elderly patient (1788). *British Journal of Psychiatry* 160: 416–8.

Förstl, H., Almeida, O.P., Owen, A.M., Burns, A. & Howard, R. (1991). Psychiatric, neurological and medical aspects of misidentification syndromes: A review of 260 cases. *Psychological Medicine* 21: 905–10.

Frade, P., Souza, D., Geraldes, R., Bentes, C., Pinho e Melo, T. & Maltez, J. (2013). Hallucinations and delusions in a hospitalized patient in a stroke unit: seizures or delirium? *European Psychiatry* 28: 1881.

Frankish, K. (2009). Delusions: a two–level framework. In M.R. Broome & L. Bortolotti, eds., *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford University Press;

Freeman, D. & Garety, P. A. (2006). Delusions. In E. Fisher & W. O'Donohue, eds., *Practitioners' Guide to Evidence–Based Psychotherapy*. Springer Academic.

Freud, S. (1911/2003). *The Schreber Case*, trad. Andrew Webber. Penguin Classics.

Frith, C.D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Psychology Press.

Frith, C.D. & Done, D.J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine* 19: 359–363.

Fulford, K.W.M. (1993). Thought insertion and insight: disease and illness paradigms of psychotic disorder. In M. Spitzer, F. Uehlin, M. Schwartz & C. Mundt, eds., *Phenomenology, Language, and Schizophrenia*. Springer.

Fusar–Poli, P., Howes, O., Valmaggia, L. & McGuire, P. (2008). 'Truman' signs and vulnerability to psychosis. *British Journal of Psychiatry* 193(2): 168.

Gallagher, S. (2009). Delusional realities. In M. Broome and L. Bortolotti, eds., *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford University Press.

Garety, P. & Hemsley, D. (1997). *Delusions: Investigations into the Psychology of Delusional Reasoning*. Oxford University Press.

Gerrans, P. (1999). Delusional misidentification as sub–personal disintegration. *The Monist* 82: 590–608.

Gerrans, P. (2000). Refining the explanation of Cotard's delusion. In M. Coltheart and M. Davies, eds., *Pathologies of Belief*. Blackwell.

Gerrans, P. (2001). Delusions as performance failures. *Cognitive Neuropsychiatry* 6(3): 161–173.

Gerrans, P. (2002). A one–stage explanation of the Cotard delusion. *Philosophy, Psychiatry, & Psychology* 9(1): 47–53.

Gerrans, P. (2009). Mad scientists or unreliable narrators? Dopamine dysregulation and delusion. In M. Broome and L. Bortolotti, eds., *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford University Press.

Gerrans, P. (2013). Delusional Attitudes and Default Thinking. *Mind & Language* 28(1): 83–102.

Gerrans, P. (2014). *The Measure of Madness: Philosophy of Mind, Cognitive Neuroscience, and Delusional Thought*. MIT Press.

Giosan, C., Glovsky, V. & Haslam, N. (2001). The Lay Concept of 'Mental Disorder': A Cross–Cultural Study. *Transcultural Psychiatry* 38: 317–332.

Glovksy, V. & Haslam, N. (2003). Acculturation and changing concepts of mental disorder: Brazilians in the U.S.A. *Transcultural Psychiatry* 40: 51–62.

Gold, I. & Hohwy, J. Rationality and Schizophrenic Delusion. In M. Coltheart and M. Davies, eds., *Pathologies of Belief*. Blackwell.

Gold, J. & Gold, I. (2012). The "Truman Show" delusion: Psychosis in the global village. *Cognitive Neuropsychiatry* 17(6): 455–472.

Goodman, N. (1978). *Ways of Worldmaking*. Hackett.

Goodman, N. (1983). *Fact, Fiction, and Forecast*. Harvard University Press.

Graham, G. (2010). *The Disordered Mind: An Introduction to Philosophy of Mind and Mental Illness*. Routledge.

Hacking, I. (1991). A tradition of natural kinds. *Philosophical Studies* 61: 109–126.

Hacking, I. (1995). The looping effects of human kinds. In D. Sperber, D. Premack & J.A. Premack, eds., *Causal cognition: A multidisciplinary debate*. Oxford University Press.

Hacking, I. (1999). Kind Making: The Case of Child Abuse. In *The Social Construction of What?*. Harvard University Press.

Hacking, I. (2007). Putnam's theory of natural kinds and their names is not the same as Kripke's. *Principia* 11(1): 1–24.

Hacking, I. (2007b). Kinds of People: Moving Targets. *Proceedings of the British Academy* 151: 285–318.

Hamilton, A. (2007). Against the belief model of delusion. In M.C. Chung, K.M.W. Fulford & G. Graham, eds., *Reconceiving Schizophrenia*. Oxford University Press.

Haslam, N. (2002). Kinds of kinds: A conceptual taxonomy of psychiatric categories. *Philosophy, Psychiatry, & Psychology* 9(3): 203–217.

Haslam, N. (2003). Categorical vs. dimensional models of mental disorder: The taxometric evidence. *Australian and New Zealand Journal of Psychiatry* 37: 696–704.

Haslam, N. (2005). Dimensions of folk psychiatry. *Review of General Psychology* 9: 35–47.

Haslam, N. (2014). Natural Kinds in Psychiatry: Conceptually Implausible, Empirically Questionable, and Stigmatizing. In H. Kincaid & J. S. Sullevin, eds., *Classifying Psychopathology: Mental Kinds and Natural Kinds*. MIT Press.

Haslam, N., Ban, L. & Kaufmann, L. (2007). Lay conceptions of mental disorder: The folk psychiatry model. *Australian Psychologist* 42(2): 129–137.

Haslam, N. & Giosan, C. (2002). The lay concept of "mental disorder" among American undergraduates. *Journal of Clinical Psychology*, 58, 479–485.

Held, R. (1961). Exposure–history as a factor in maintaining stability of perception and coordination. *The Journal of Nervous and Mental Disease* 132: 26–32.

Hohwy, J. (2004). Top–down and bottom–up in delusion formation. *Philosophy, Psychiatry, & Psychology* 11(1): 65–70.

Hohwy, J. (2013). Delusions, Illusions and Inference Under Uncertainty. *Mind & Language* 28(1): 57-71.

Hohwy, J. & Rosenberg, R. (2005). Unusual experiences, reality testing and delusions of alien control. *Mind & Language* 20(2): 141–162.

Horwitz, A.V. & Wakefield, J.C. (2012). *The Loss of Sadness. How Psychiatry Transformed Normal Sorrow into Depressive Disorder*. Oxford University Press.

Howson, C. & Urbach, P. (1993). *Scientific Reasoning: The Bayesian Approach*. Open Court.

Hull, D. (1978). A matter of individuality. *Philosophy of Science* 45: 335–360.

James, W. (1988). *The Listening Ebony: Moral Knowledge, Religion and Power among the Uduk of Sudan*. Clarendon Press.

Jaspers, K. (1913/1963). *General Psychopathology*, trad. J. Hoenig & M.W. Hamilton. Manchester University Press.

Jeannerod, M. (2003). Action Monitoring and Forward Control of Movements. In M. Arbib, ed., *The Handbook of Brain Theory and Neural Networks*. Second Edition. MIT Press.

Kanner, L. (1935). *Child Psychiatry*. Charles C. Thomas.

Kendler, K., Zachar, P. & Craver, C. (2011). What kinds of things are psychiatric disorders. *Psychological Medicine* 41: 1143–1150.

Krabbendam, L., Myin–Germeys, I., Bak, M., & van OS, J. (2005). Explaining transitions over the hypothesized psychosis continuum. *Australian and New Zealand Journal of Psychiatry* 39(3): 180–186.

Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman, eds., *Semantics of Natural Language*. Reidel.

Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*. Chicago University Press.

Kumazaki T. (2011). What is a 'mood–congruent' delusion? History and conceptual problems. *History of Psychiatry* 87: 315–31.

Jackson, A., Cavanagh, J. & Scott, J. (2003). A systematic review of manic and depressive prodromes. *Journal of Affective Disorders* 74(3): 209–217.

Lam, D. & Wong, G. (2005). Prodromes, coping strategies and psychological interventions in bipolar disorders. *Clinical Psychology Review* 25(8): 1028–1042.

Langdon, R. & Coltheart, M. (2000). The cognitive neuropsychology of delusions. In M. Coltheart & M. Davies, eds., *Pathologies of Belief*. Blackwell.

Leeser, J. & O'Donohue, W. (1999). What is a delusion? Epistemological dimensions. *Journal of Abnormal Psychology* 108: 687–694.

Lewis, D.K. (1982). Logic for Equivocators. *Noûs* 16(3): 431–441.

Livesley, W. (2003). Diagnostic dilemmas in classifying personality disorder. In K. Phillips, M. First & H. Pincus, eds., *Advancing DSM: Dilemmas in Psychiatric Diagnosis*, American Psychiatric Association.

Maher, B.A. (1999). Anomalous experience in everyday life: Its significance for psychopathology. *The Monist* 82: 547–70.

Maher, B.A. (2001). Delusions. In Patricia B. Sutker & Henry E. Adams, eds., *Comprehensive Handbook of Psychopathology*. Kluwer Academic.

Manschreck, T.C. (1979). The assessment of paranoid features. *Comprehensive Psychiatry* 20: 370–377.

McKay, R., Langdon, R. &): Coltheart, M. (2009). "Sleights of Mind": Delusions and Self–deception. In Philosophy, Psychiatry, & PsychologyT. Bayne and J. Fernández, eds., *Delusion and Self–Deception*. Psychology Press.

Miyazono, K. Bortolotti, L. (forthcoming). The causal role argument against doxasticism about delusions. *AVANT*.

Mojtabai, R. (1994). Fregoli syndrome. *Australian & New Zealand Journal of Psychiatry* 28(3): 458–62.

Mojtabai, R. & Nicholson, R. (1995). Interrater reliability of ratings of delusions and bizarre delusions. *American Journal of Psychiatry* 152: 1804–8.

Mullen, R. & Gillett, G. (2014). Delusions: A Different Kind of Belief? *Philosophy, Psychiatry, & Psychology* 21(1) :27-37.

Murphy, D. (2006). *Psychiatry in the Scientific Image*. MIT Press.

Murphy, D. (2012) The Folk Epistemology of Delusions. *Neuroethics* 5(1): 19–22.

Murphy, D. (2013). Delusions, Modernist Epistem–ology and Irrational Belief. *Mind & Language* 28(1): 113–124.

Murphy, D. (2014). Natural Kinds in Folk Psychology and in Psychiatry. In H. Kincaid & J. S. Sullevin, eds., *Classifying Psychopathology: Mental Kinds and Natural Kinds*. MIT Press.

Nasar, S. (1998). *A Beautiful Mind*. Simon & Schuster.

Nijinsky, V. (1995). *The Diary of Vaslav Nijinsky*, ed. J. Acocella. Farrar, Straus & Giroux.

Nozick, R. (1993). *The Nature of Rationality*. Princeton University Press.

O'Dwyer, J.M. (1990). Coexistence of the Capgras and de Cl erambault's syndromes. *British Journal of Psychiatry* 156: 575–77.

Okasha, S. (2002). Darwinian Metaphysics: Species and the Question of Essentialism. *Synthese* 131: 191–213.

Page, S. (2006). Mind–independence disambiguated: separating the meat from the straw in the realism/anti–realism debate. *Ratio* 19(3): 321–335.

Parnas, J. & Sass, L.A. (2001). Self, solipsism, and schizophrenic delusions. *Philosophy, Psychiatry, & Psychology* 8(2–3): 101–120.

Perceval, J.T. (1840). Narrative on the Treatment Experienced by a Gentleman during a State of Mental Derrangement. Effingham Wilson.

Price, H.H. (1960/1969). *Belief*. George Allen and Unwin.

Putnam, H. (1975). The meaning of 'meaning'. In *Mind Language, and Reality*. Cambridge University Press.

Quine, W.V. & Ullian, J. (1970). *The Web of Belief*. Random House.

Radden, J. (2011). *On Delusion*. Routledge.

Reid, I., Young, A.W. & Hellawell, D.J. (1993). Voice recognition impairment in a blind Capgras patient. *Behavioural Neurology* 6(4): 225–8.

Reimer, M. (2010). Only a Philosopher or a Madman: Impractical Delusions in Philosophy and Psychiatry. *Philosophy, Psychiatry, & Psychology* 17(4): 315–328.

Rodrigues, A., Banzato, C., Dantas, C. & Dalgalarrondo, P. (2013). Capgras Syndrome. In Fred. R. Volkmar, ed., *Encyclopedia of Autism Spectrum Disorders*. Springer.

Rose, D., Buckwalter, W. & Turri, J. (2014). When words speak louder than actions: delusion, belief and the power of assertion. *Australasian Journal of Philosophy* 92(4): 683–700.

Ryle, G. (1949). *The Concept of Mind*. Hutchinson.

Samuels, R. (2009). Delusions as a Natural Kind. In M. Broome & L. Bortolotti, eds., *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford University Press.

Samuels, R. & Ferreira, M. (2010). Why Don't Concepts Constitute a Natural Kind? *Behavioral and Brain Sciences* 33(2–3): 222–223.

Sass, L.A. (1992). *Madness and Modernism*. Harvard University Press.

Sass, L.A. (1994). *The Paradoxes of Delusion: Wittgenstein, Schreber, and the Schizophrenic Mind*. Cornell University Press.

Sass, L.A. (2004). Some reflections on the (analytic) philosophical approach to delusion. *Philosophy, Psychiatry, & Psychology* 11(1): 71–80.

Schreber, D.P. (1903/2000). *Memoirs of My Nervous Illness*, trad. I. Macalpine & R.A. Hunter. New York Review of Books.

Schwitzgebel, E. (2001). In–between believing. *Philosophical Quarterly* 51: 76–82.

Schwitzgebel, E. (2012). Mad belief. *Neuroethics* 5(1): 13–17.

Sechehaye, M. (1994). *Autobiography of a Schizophrenic Girl*, trad. Grace Rubin–Rabson. Meridian.

Shimizu, M., Kubota, Y., Toichi, M. & Baba, H. (2007). *Folie à deux* and Shared Psychotic Disorder. *Current Psychiatry Reports* 9: 200–5.

Škodlar, B., Dernovšek, M.Z Kocmur, M. (2008). Psychopathology of Schizophrenia in Ljubljana (Slovenia) From 1881 To 2000: Changes in the Content of Delusions in Schizophrenia Patients Related To Various Sociopolitical, Technical and Scientific Changes. *International Journal of Social Psychiatry* 54(2): 101–111.

Solomon, M. (2000). "Twenty–Five Heads under One Hat: " Quick–Change in the 1890s. In Vivian C. Sobchack, ed., Meta Morphing: Visual Transformation and the Culture of Quick–change. University of Minnesota Press.

Somerfield, D. (1999). Capgras syndrome and animals. International Journal of Geriatric Psychiatry 14(10): 893–894.

Spence, S.A., Hirsch, S.R. Brooks, D.K. & Grasby, P.M. (1998). Prefrontal cortex activity in people with schizophrenia and control subjects. Evidence from positron emission tomography for remission of 'hypofrontality' with recovery from acute schizophrenia. *British Journal of Psychiatry* 172: 316– 323.

Sperber, D. (1980). Is Symbolic Thought Pre–rational? In M. Foster & S. Brandes, eds., *Symbol as Sense*. Academic Press.

Sperber, D. (1982). *On Anthropological Knowledge*. Cambridge University Press.

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Blackwell.

Spitzer, M. (1990). On defining delusions. *Comprehensive Psychiatry* 31: 377–397.

Stephens, G.L. & Graham, G. (2000). *When Self–Consciousness Breaks*. MIT Press.

Stephens, G.L. & Graham, G. (2004). Reconceiving delusions. *International Review of Psychiatry* 16(3): 236–241.

Stephens, G.L. & Graham, G. (2007). The Delusional Stance. In M.C. Chung, K.W.M. Fulford & G. Graham, eds., *Reconceiving Schizophrenia*. Oxford University Press.

Stich, S.P. (1979). Do Animals Have Beliefs? *Australasian Journal of Philosophy* 57: 15-28.

Stich, S.P. (1990). *The Fragmentation of Reason*. MIT Press.

Stompe, T., Friedman, A., Ortwein, G., Strobl, R., Chaudhry, H.R., Najam, N. & Chaudhry, M.R. (1999). Comparison of delusions among schizophrenics in Austria and in Pakistan. *Psychopathology* 32(5): 225–234.

Stone, T. & Young, A.W. (1997). Delusions and brain injury: The philosophy and psychology of belief. *Mind & Language* 12: 327–64.

Szasz, T. (1961). *The Myth of Mental Illness: Foundations of a Theory of Personal Conduct*. Harper & Row.

Tateyama, M., Asai, M., Hashimoto, M., Bartels, M. & Kasper, S. (1998). Transcultural study of schizophrenic delusions. Tokyo versus Vienna and Tubingen (Germany). *Psychopathology* 31(2): 59–68.

Tranel, D., Damasio, H. & Damasio, A.R. (1995). Double dissociation between overt and covert recognition. *Journal of Cognitive Neuroscience* 7: 425–32.

Vallar, G., & Ronchi, R. (2009). Somatoparaphrenia: A body delusion. A review of the neuropsychological literature. *Experimental Brain Research* 192: 533–551.

van Os, J., Linscott, R.J., Myin–Germeys, I., Delespaul, P. & Krabbendam, L. (2009). A systematic review and meta–analysis of the psychosis continuum: evidence for a psychosis proneness–persistence–impairment model of psychotic disorder. *Psychological Medicine* 39: 179–195.

van Os, J. & Kapur, S. (2009). Schizophrenia. *The Lancet* 374: 635–645.

van't Veer–Tazelaar, P.J., van Marwijk, H.W., van Oppen, P., van Hout, H.P., van der Horst, H.E., Cuijpers, P., Smit, F. & Beekman, A.T. (2009). Stepped–care prevention of anxiety and depression in late life: a randomized controlled trial. *Archives of General Psychiatry* 66(3): 297–304.

Wallis, G. (1986). Nature of the misidentified in the Capgras syndrome. *Bibliotheca Psychiatrica* 164: 40–48.

Wang, E.W. (2014). Perdition Days: On Experiencing Psychosis. *The Toast*. URL = <http://the-toast.net/2014/06/25/perdition-days-experiencing-psychosis/>

Wakefield, J.C. (1992). Disorder as harmful dysfunction: a conceptual critique of DSM–III–R's definition of mental disorder. *Psychological Review* 99: 232–247.

Wakefield, J.C. (2004). The myth of open concepts: Meehl's analysis of construct meaning versus black box essentialism. *Applied & Preventive Psychology* 11(1): 77–82.

Weinstein, E.A. and Kahn, R.L. (1955) *Denial of Illness*. Charles C. Thomas.

Wessely, S., Buchanan, A., Reed, A., Cutting, J., Everitt, B., Garety, P. & Taylor, P.J. (1993). Acting on delusions. I: Prevalence. *British Journal of Psychiatry* 163: 69–76.

Widiger, T., Livesley, W. & Clark, L. (2009). An integrative dimensional classification of personality disorder. *Psychological Assessment* 21(3): 243–255.

Widiger, T., & Sanderson, C. (1995). Toward a dimensional model of personality disorders. In W. Livesley, ed., *The DSM–IV Personality Disorders*. Guilford.

Wittgenstein, L. (1969). *On Certainty*. Ed. G.E.M. Anscombe, G.H. Wright & D.Paul. Trad. G.E.M. Anscombe & D. Paul. Blackwell.

Young, A.W. (1999). Delusions. *The Monist* 82(4): 571–589.

Young, A.W. (2000). Wondrous strange: The neuropsychology of abnormal beliefs. In M. Coltheart and M. Davies, eds., *Pathologies of Belief*. Blackwell.

Young, A.W. & Leafhead, K. (1996). Betwixt life and death: Case studies of the Cotard delusion. In P. Halligan & J. Marshall, eds., *Method in Madness*. Psychology Press.

Zachar, P. (2000). Psychiatric disorders are not natural kinds. *Philosophy, Psychiatry, & Psychology* 7(3): 167–182.

Zachar, P. (2014). Beyond natural kinds: Toward a relevant psychiatric taxonomy. In H. Kincaid & J. S. Sullevin, eds., *Classifying Psychopathology: Mental Kinds and Natural Kinds*. MIT Press.

Zachar, P. (2014b). *A Metaphysics of Psychopathology*. MIT Press.

Zimbardo, P.G., Andersen, S.M. & Kabat, L.G. (1981). Induced hearing deficit generates experimental paranoia. *Science* 212: 1529–31.