DELUSIONS, ACCEPTANCES, AND COGNITIVE FEELINGS

BY RICHARD DUB

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Philosophy

Written under the direction of

Brian McLaughlin

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2013

ABSTRACT OF THE DISSERTATION

Delusions, Acceptances, and Cognitive Feelings

by Richard Dub

Dissertation Director: Brian McLaughlin

Psychopathological delusions, such as the Capgras delusion, the Cotard delusion, and the florid delusions that accompany schizophrenia, have a number of features that are curiously difficult to explain. Delusions are resistant to counterevidence and impervious to counterargument. They are theoretically, affectively, and behaviorally circumscribed; delusional individuals tend not to act on their delusions or draw appropriate inferences from the content of their delusions. Delusional individuals are occasionally able to distinguish their delusions from other beliefs, sometimes speaking of their "delusional reality." I argue that these features support non-doxasticism about delusions. Non-doxasticism is the thesis that, contrary to appearances, delusions are not beliefs at all. After developing the prospects for non-doxasticism, I offer a novel non-doxasticist cognitive model. Delusions are pathological acceptances that are caused by powerful and aberrant cognitive feelings.

Acknowledgements

As I have worked through the issues in this dissertation, my committee members have constantly been invaluable wellsprings of inspiration and guidance. I owe much to them. Andy Egan has been a wonderful source of incisive comments about the nature of belief-like mental states. Conversation with him is always immensely fruitful; his suggestions on previous drafts have been top-notch. Stephen Stich's writings have always been a large influence on my thought and they have been a large inspiration of the account of cognition that I provide herein. His comments on and criticisms of earlier versions of my proposal have have prompted many major revisions to my thought. Louis Sass has been my window into an academic field other than my own. I have learned much about delusions, phenomenology, and clinical psychology from him, and this dissertation has benefited greatly from his help.

I have further benefited from conversations with graduate students in philosophy and psychology at Rutgers and in the New York area, including (but certainly not limited to) Josh Armstrong, Tom Donaldson, E. J. Green, Michael Johnson, Ben Levinstein, Zak Miller, Lisa Miracchi, Jennifer Nado, David Rose, Karen Shanton, and Jennifer Wang. Frankie Egan and Robert Matthews have additionally provided indispensable commentary. I would also like to thank my audiences at the Rutgers Center for Cognitive Science, the British Postgraduate Philosophy Association, and the Institute of Cognitive Science at the Universtität Osnabrück. Extra special thanks must be directed to Elizabeth Pienkos for her tireless support, her patience with my travails, her help in reading drafts, and her shared commitment to my goals.

Of course, the lion's share of the gratitude must go to my dissertation director,

Brian McLaughlin. Brian has been remarkably generous with his time and dedication; his encouragement, suggestions, and mentorship have all been absolutely crucial.

To all of these people, I owe my thanks.

Dedication

To Liz.

Table of Contents

Chapter 1
Introduction

1.1 Overview

This dissertation provides a model of psychotic delusions in which delusions are not beliefs.

The model responds to two different debates in the literature on delusions that are not often brought together. The first debate concerns the *etiology*—the causal origin—of delusions. How does a person suffering from a psychosis come to form a delusional belief? Why do some individuals come to believe that their spouse has been replaced with imposter? There are two main types of answer to these questions. Each makes a different claim about the number of cognitive mechanisms that need to be broken in order for a delusion to develop. ***One-factor*** theories hold that we only need to appeal to a single impairment. Delusions are normal responses to an anomalous, pathological experience. ***Two-factor*** theories, on the other hand, hold that anomalous experience is not in itself sufficient to produce a delusion. We also need to explain how the experience leads to a delusional belief, and why that belief is not immediately rejected for being bizarre. This requires us to posit a second pathology: an impairment or a pathological bias in whatever mechanism is involved in belief fixation. The second debate concerns the *characterization* of delusions. It is commonly assumed that delusions are beliefs of a certain sort. The one-factor/two-factor debate typically makes this assumption: it is typically presented as a debate about how delusional beliefs are formed. Theories that claim that delusions are beliefs are ***doxasticist*** theories. Lately, some philosophers and psychologists have argued that delusions don't behave as one would expect beliefs to behave. Hence, they conclude, delusions are not beliefs. Rather, they are some other sort of mental state.

Theories that make this claim are known as ***non-doxasticist*** theories.

I defend a *non-doxasticist, one-factor* model of delusions. In brief: a delusion is not a belief, but a distinct sort of mental state that I will call a *acceptance*, and it is formed on the basis of a powerful and anomalous *cognitive feeling*.

Acceptances are a type of mental state distinct from beliefs. A person can accept that he is being followed, for instance, without genuinely believing that he is being followed. Other philosophers have proposed similar mental states that they call 'avowals' or 'opinions'. On the model I propose, cognitive feelings have an important influence on acceptances. They can cause one to adopt an acceptance non-volitionally.

Cognitive feelings are a type of mental state that are only lately being explored by cognitive science.[1] Examples of cognitive feelings include the feeling of confidence, the feeling of familiarity, or the feeling that one is being followed.

I have tried to keep the account free of assumptions that would render it palatable to only readers of a particular theoretical stripe. The account is unabashedly naturalist and empirical, but beyond these minimal commitments, it is intended to be compatible with most dispositional, functional, representational, or computational theories of mentation.

1.2 Goals

Why is it interesting or valuable to develop a model of delusions? Throughout the writing of this dissertation, I have been guided by (at least) three reasons for thinking that this is an important subject of inquiry. Each has given me a different goal. These goals range from being relatively narrow to having fairly wide-reaching consequences for philosophy.

1. I argue for a particular conception of the pathology suffered by delusional individuals.

[1] See, for example, Clore and Gaspar (1992, 2000), Ratcliffe (2005, 2008), and McLaughlin (2009).

Quite regardless of any wider implications that this project might have for philosophy or psychology, psychotic delusions are fascinating in their own right and worthy of investigation in themselves. Moreover, there are practical reasons to give an accurate account of delusions. If the account that I offer is correct, then clinical implications would likely follow. The account could very well prove useful for therapeutic or diagnostic purposes. (However, I won't draw any conclusions about the clinical ramifications of the model in this dissertation.)

2. I argue for a particular conception of the cognitive architecture of the normal, non-pathological human mind.

   Breakages are often very revealing to the reverse-engineer. When computer programs crash, we can learn something about the structure of the program and its underlying algorithms. When the mind crashes, we can learn something about the structure of thought. For this reason, philosophers of mind have been keenly interested in developments in abnormal and clinical psychology and in cognitive neuroscience. Lesion and deficit studies performed by cognitive neuroscientists, for instance, tell us about the sorts of psychological impairments that an individual suffers when a part of the brain has undergone damage, and this yields information about the cognitive tasks that are enabled by neural structures in that location.

   I propose that the best explanation of delusions incorporates acceptances and cognitive feelings as mental states. These states are also operative in everyday non-pathological, non-delusive cognition. Thus, the theory has implications that are much wider that reach beyond the realm of psychopathology.

3. I argue for a *lush*, rather than *sparse*, conception of mental states.

   There's a tendency among philosophers to admit only a small stock of mental states or attitudes when giving psychological explanations. According to many, "belief-desire psychology" is successful in everyday life and is still the language in which psychologists conduct their explanations, and this is enough to vindicate realism about beliefs and desires. Depending on how one chooses to

understand the phrase "belief-desire psychology", this might mean that beliefs and desires are only two of many psychological states that we might call upon in giving psychological explanations. However, in practice, many philosophers who hold fast to belief-desire psychology use beliefs and desires in pretty much all explanations, and are extremely reluctant to call upon other attitudes. Those who think we can make do with just a small stock of mental states in our explanations think that the mind is sparse. Should we think that the mind is sparse rather than lush?

Doxasticism is the thesis that delusions are beliefs and not some other sort of mental state. I see this as representative of the conservatism that underwrites a sparse conception of mental states. The first half of the dissertation adjudicates the arguments between the doxasticists and non-doxasticists, and many of the arguments offered are not simply about delusions: they are general enough to be seen as arguments for and against a lush or a sparse conception of mental states. The implications of these debates are far-reaching. What distinguishes beliefs and desires from other propositional attitudes? Should we expect all propositional attitudes to have a direction of fit? Are beliefs and desires essential to cognition? Are these simply terminological debates about the extension of 'belief' and 'desire'? Chapter 5 addresses these issues directly.

1.3 Chapter Summaries

The structure of the dissertation is this: the first chapter following this introduction sets up the problem to be addressed; the next two chapters attack competing attempts that have been made by non-doxasticists to address the problem; the chapter afterward sets up the positive proposal; the final chapter gives the positive proposal itself. Chapters 2, 4, and 6 are most explicitly about delusions. Reader coming at this dissertation from psychology or psychiatry rather than philosophy might find their interests best served by focussing on these chapters.

I have attempted to leave signposts and summaries throughout each chapter to

help the reader understand where he or she is in the argument. Nonetheless, I invite the reader to refer to the The Table of Contents often, as the section headings can be used as a roadmap.

Chapter 2: Delusion as Non-Belief introduces to the reader descriptions and characterizations of delusion that have been given in the literature. I show that such characterizations typically presume doxasticism, but that this presumption is unwarranted. I then describe the features that are in need of explanation and that have motivated philosophers to provide non-doxasticist accounts.

Chapter 3: Conceptual Arguments and Rationality deals with the arguments that are most often made for non-doxasticism. According to these arguments, beliefs are necessarily rational, but delusions are not rational, so delusions cannot be beliefs. I argue that these arguments are unsuccessful, and that there is no rationality requirement on belief. If non-doxasticism is to be successful, non-doxasticists should give explanatory theories rather than attempt a priori arguments that call upon conceptual requirements of belief.

Chapter 4: Unsuccessful Non-Doxasticist Theories considers non-doxasticist proposals that I find wanting. Major theories attacked include those of Currie, Egan, Schwitzgebel, and Campbell. A major theme of this chapter will be that it is easy for an allegedly non-doxasticist account to actually be a terminological variant of a doxasticist account.

Chapter 5: Concepts of Belief attempts to assuage the critic who thinks that the concept of belief is so broad that it will necessarily cover delusion, and that non-doxasticism is too revisionary. I diagnose a reason for thinking that delusions must be beliefs and argue against it, and then show that the notion of acceptance is not as revisionary as it might at first seem, for folk psychology recognizes a rough distinction between beliefs and acceptances.

Chapter 6: A Positive Proposal presents the model of delusion that has been promised throughout. Delusions are acceptances based on cognitive feelings. Most

of the chapter is dedicated to characterizing the notion of acceptance. At the end, I argue for a particular relation between acceptance and cognitive feeling, and explain how my theory explains the features of delusion with which other theories have such trouble.

Chapter 2
Delusion as Non-Belief

Philosophers in need of colorful ways to illustrate irrational belief have long looked to the madman to supply them with examples. Descartes opens his First Meditation by enjoining the reader to consider "madmen whose brains are so damaged by the persistent vapours of melancholia that they firmly maintain they are kings when they are paupers, or say they are dressed in purple when they are naked, or that their heads are made of earthenware, or that they are pumpkins, or made of glass" (1641/1986). These examples are in fact more prosaic than many of the truly bizarre delusions that are relatively common among the mentally ill. Sufferers of the Capgras delusion might claim that their spouse or a loved one has been replaced with an imposter. Sufferers of the Cotard delusion hold fast to an unshakeable conviction that they are dead. Many schizophrenics suffer from thought insertion, and say that they are the thinker of someone else's thoughts, and others suffer from delusions of reference in which ordinary objects take on all manner of dark and personal import, such as signifying that they are about to be murdered. Delusional subjects usually cannot be argued out of their fantastic stories, even in the face of overwhelming evidence to the contrary. It is incredible to many of us that delusional subjects say what they do and act as they do. Delusions are surprising and eerie, but therefore intriguing. What features of the delusional individual's cognitive architecture could be responsible?

The theorist interested in explaining delusions is faced with questions about their constitution and their provenance—that is, questions about their ontology and their etiology. What are delusions? How are they formed?

Traditionally, delusions have been thought to be false beliefs formed on the basis

of abnormally irrational reasoning. There are reasons to be skeptical of this characterization. Because of the unusual features of delusions, many philosophers and psychologists have found it problematic to interpret delusions as irrational beliefs. Delusions are like prototypical beliefs in some respects but not in others; categorizing them as beliefs is uncomfortable at best.

Suppose you form the belief that a doppelgänger has taken the place of your spouse. How should I expect you to act? I would think that you would contact the authorities, hide from the alleged imposter, try to find your real spouse, and try to figure out what led to this outlandish situation. The Capgras patient typically does none of these things. Similarly, one would expect that a person who believes that the world will end tomorrow would try to get their affairs in order, try to stop the event, or at the very least show fear. The schizophrenic suffering from delusions of catastrophe might do none of these things. Patients might claim that their food has been poisoned by spies at the very same moment that they happily tuck into it (Sass 1994). Bleuler (1950) remarks that those who claim to be dogs don't bark like dogs. The odd fact in need of explanation is that, in most cases, a delusional patient's behavior seems to outright belie what the patient asserts. It appears that they are mouthing the words without really believing what they are saying. Psychiatrists have long written of their doubts that their delusional clients report genuine beliefs.

Could a person really believe that her food is poisoned and yet still continue to eat? Her behavior could be explained if she had other aberrant beliefs and desires, such as a desire to die, or a belief that poison is harmless. However, patients usually seem to lack these sorts of intermediary mental states. Patients will sometimes confabulate extremely unlikely and irrational justifications for their delusions, but the explanations are usually pretty clearly made-up on the spot, and they are just as problematic as the original delusion, for they do not manifest themselves elsewhere in the patient's reasoning or behavior. Chris Frith and Eve Johnstone (2003) recall a patient of Alan Baddesky's who claimed to be a Russian chess Grand Master. When asked why he didn't speak in Russian given that he claimed he was Russian, he replied that he may have been "hypnotized to forget things like the fact that [he]

could speak Russian" (Frith and Johnstone 2003, p.152). Nothing else about his behavior suggested he believed he was a target for malevolent mesmerists. He possibly would have confabulated another justification if prodded, though patients who are constantly asked to justify their delusions are often uncomfortable with the positions to which they are pushed and will stop responding.

Similarly, in an interview with documentarian Errol Morris (2010), the neurologist V. S. Ramachandran describes a patient with a paralyzed arm who is afflicted with anosagnosia (the refusal to accept that one is suffering from an injury) and asomatagnosia (the refusal to accept that one's own body part is actually one's own) as follows:

> "I grabbed her left arm and I said, 'Whose arm is this?' She said, 'That's my mother's arm.' ...And I said, 'Well, if that's your mother's arm, where's your mother?' And she looks around, completely perplexed, and she said, 'Well, she's hiding under the table.' I said, 'Please touch your nose with your left hand.' She immediately takes her right hand, goes and reaches for the left hand, raising it, passively raising it, right? Using it as a tool to touch my nose or touch her nose. ...What does this imply? She claims her left arm is not paralyzed, right? Why does she spontaneously reach for it and grab her left arm with her right hand and take her left hand to her nose? That means she knows it is paralyzed at some level. ...She's just now told me that it's not her left arm, it is her mother's arm, so why is she pulling up her mother's arm and pointing it at my nose?" (Morris 2010)

It's difficult to know how to best characterize this patient's condition. Ramachandran's claim that his patient knows that her arm is her arm "at some level" clearly calls out for further explanation. The unlikely comportment of those who suffer from delusions has long puzzled psychiatrists. Some have concluded that the delusional patient should not be understood as genuinely reporting beliefs. Ramachandran's patient can't *really* believe that her arm belongs to her mother, and the patient who says her food is poisoned can't *really* believe it. They're just mouthing the words. The apparent assertions of the madman might be "senseless ravings" beyond understanding (Jaspers 1963, p.577). Berrios claims that the reports of the delusive are "empty speech acts that disguise themselves as beliefs" (Berrios 1991, p.8). Alternatively, some psychiatrists opt to understand the content of a delusional subject's

utterances metaphorically (Laing 1969).

These sorts of conclusions have commanded the attention of philosophers. The unusual behavior of delusional subjects appears to be in conflict with a functionalist conception of belief. It is largely accepted that for a large class of mental states that includes propositional attitudes such as beliefs and desires, each token mental state is the type of state that it is in virtue of playing a particular functional role in an agent's cognitive economy. In other words, a mental state is a belief-that-*p* only if it functions like a belief-that-*p*.[1] Delusions apparently fail to play the functional role of beliefs. The role played by delusions diverges from the role played by prototypical beliefs in the following ways, for instance:

- Delusions are usually poorly integrated and are often outright inconsistent with other beliefs.
- Delusional patients are not sensitive to evidence and fail to revise their delusions when confronted with evidence to the contrary.
- Delusional patients fail to act appropriately on their delusions and fail to use them in practical reasoning.
- Delusional patients fail to exhibit the expected sorts of emotion or affect in response to their delusion.
- Delusional patients often exhibit a metacognitive awareness of their irrationality; they know that they ought not have the delusion and often take something like a "sarcastic" attitude toward their delusion, accompanying their reports with a wry smile (Sass 1994, Young 1999, Gallagher 2009).

These data have led philosophers and psychologists to argue for a *non-doxastic conception of delusions*. Non-doxasticism is often described in a way that is pithy and punchy at the expense of being wholly precise: it holds that *delusions aren't beliefs*.

[1] Throughout this dissertation, I assume that beliefs and delusions are attitudes with propositional content. For evidence that beliefs have propositional contents, consult chapter 3 of Stich (1983).

Usually, the non-doxasticist will also offer a positive account that explains what delusions are. For example, Currie (2000) holds that delusions are misidentified imaginings; Egan (2009) holds that they are instances of a state intermediate between belief and imagination.

A major goal of this dissertation is to appraise arguments for and against non-doxasticism, and to offer a particular non-doxasticist model of delusions—one that answers the questions about ontology and etiology. On the model I will present and defend, delusions are *acceptances* formed on the basis of pathological *cognitive feelings*.

First things first. What are the basic building blocks upon which to build a non-doxasticist model? Clinicians often take non-doxasticism to be a surprising and non-standard thesis. It is natural to describe delusions as beliefs. For instance, *The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*, by far the most widely used diagnostic text in psychiatric practice, defines delusions as false beliefs not normally accepted by one's subculture that are based on incorrect inference about external reality and that persist despite counterevidence. Why should we doubt this definition? In this initial chapter, I'll present some traditional characterizations of delusions, and then, in counterpoint, describe the more bewildering features of delusions that have motivated non-doxasticists to resist these traditional characterizations. Whether these data actually do support non-doxasticism will be a subject for future chapters.

2.1 Traditional Characterizations of Delusions

Delusions have long been the archetypal indicator of madness and psychosis. They are one of the "reality distortion" symptoms; to be deluded is to have lost one's grip on the world.

This description obviously does not do a great job of telling us what delusions are. Can we do better? There are two ways of getting clearer on our subject matter: extensionally and intentionally. To give an extensional characterization of delusions,

we list the various sorts of delusions that there are; to give an intensional characterization, we specify the properties that make a mental state a delusion.

Neither of these tasks is easily accomplished.

2.1.1 Extensional Characterizations

Delusions can be categorized along a number of different dimensions, but they are most often categorized by their content. Major types include misidentification delusions such as the Capgras delusion or the Frégoli delusion (the delusion that strangers are actually familiar individuals in disguise), delusions of grandiosity, persecutory delusions, delusions of reference, religious delusions, delusional jealousy, delusions of depersonalization, sexual delusions, and so on.

In addition to categorizing delusions by their content, a number of other important distinctions can be drawn. Delusional syndromes can be *monothematic* (the subject is deluded about a single theme or topic, as in Capgras syndrome), or they can be *polythematic*/*florid* (about many different topics that are unrelated to one another). They can be *organic* (associated with localized brain injuries), or *functional* (not associated with any particular type of localized brain injury). Some delusions are said to be *mood-congruent*: they can be understood as continuous with, and partly originating in, aspects of the subject's mood. Depressive delusions or delusions of grandeur, for instance, can be understood as an expression of an individual's mood disorder. Others are *mood-incongruent*. Some delusions are *circumscribed*: they do not tend to affect behavior and other mental states. Delusions that are not circumscribed are *elaborated*. Delusions can be *bizarre* in their content, or relatively *mundane*. Some generate spontaneous confabulations; others only produce confabulations in the subject when provoked; others do not generate confabulations at all. Delusions are associated with an enormous variety of psychopathologies, including schizophreniform disorder, schizoaffective disorder, delusional disorder, dementia, mania, and certain mood disorders such as major depression.

Delusions are a particularly important diagnostic criterion for schizophrenia. Ones particularly associated with schizophrenia include:

THOUGHT BROADCASTING: the patient claims everyone can hear his or her thoughts.

PASSIVITY PHENOMENA: the patient claims that thoughts are implanted in his or her head through radio waves.

THOUGHT INSERTION: the patient claims they are thinking someone else's thoughts.

DELUSIONS OF REFERENCE: the patient claims that various acts and objects are messages to him or her.

DELUSIONS OF CATASTROPHE: the patient claims the world is coming to an end.

DELUSIONS OF DEREALIZATION: the patient claims the world isn't real.

The particular expressions of these delusions are often determined by one's culture. People suffering from delusions of thought insertion and thought broadcasting will often invoke the internet; it used to be that delusional people stated that thoughts were transmitted by radio waves. In a famous account from the late 1700's, James Tilly Matthews maintained that spies skilled in "pneumatic chemistry" could influence thoughts by manipulating magnetic fields using an "Air Loom" (Carpenter 1989).

The huge variety in types of delusion creates a problem for the theorist interested in giving a theory of delusion. Perhaps it doesn't make sense to talk about delusions as a unified group. It could be that the sorts of states we call 'delusions' are unmanageably heterogenous. Why think that there is *a* theory that can cover all these instances? Worse still, the concept doesn't pick out an obviously unproblematic class; even if all the above are uncontentiously considered delusions, there are other states that are possibly delusions but possibly not. Psychiatric symptoms such as unresisted obsessions are arguably delusions. Coltheart et al. (2011) claim that, of the reality distortion symptoms, hallucinations are straightforwardly distinct from delusions: hallucinations are often thought of as false perceptions, whereas delusions are thought of as false beliefs. However, even hallucinations are suspect; some phenomenologists write about "delusional perceptions" (Fuchs 2005). Moreover, an awful lot of mental states that are not normally considered pathological seem like

delusions. Consider superstitions, religious beliefs, self-deceptions, strong political beliefs, beliefs about one's own capabilities and uniqueness, and so on. There are questions about whether cultural acceptability plays a role in the categorization of a mental state as a delusion (Radden 2010, Murphy 2013), and whether these sorts of beliefs are points on a continuum that runs from the regular to the pathologically delusional.

The problem for the psychologist and the philosopher is knowing exactly which phenomena should be in the crosshairs. One methodological solution is to focus on a single type of delusion. Philosophers and cognitive psychologists have traditionally focussed on the Capgras delusion: given that it is often monothematic, circumscribed, and accompanied with specific neurological damage, it is an especially clean case to study. We have the outlines of a reasonably good theory of how the delusion is formed. However, it would be unfortunate to only focus on very particular theories when a more general theory of delusion could be given.

The idea that there is *a* general theory of delusion is a *working hypothesis*. There is some reason to think that a single general sort of explanation can be given, which makes this a justifiable starting point. Many of the delusions mentioned above can arise in a single individual at the same time—they can be comorbid with one another. The Capgras delusion can arise monothematically from damage to the frontal lobe, but it can also be one of many delusions that become manifest during the psychotic break of a schizophrenic. This suggests a common etiology and common explanation.

Still, there are many different sorts of delusion each with features unique to it, and perhaps some of these delusions will not be included in a general theory. For instance, the provenance of mood-congruent delusions is not as mystifying as that of mood-incongruent delusions, and although psychotic patients rarely act upon their delusions, this is not true of mood-congruent delusions such as delusional jealousy

and persecutory delusions (see Wessely et al. (1993)). Perhaps they should be handled differently. It may well turn out to be that paranoid delusions are irrational beliefs, but schizophrenic delusions of reference are not beliefs at all. In this dissertation, I'll focus on organic monothematic delusions such as the Capgras delusion, and the florid delusions that tend to accompany schizophrenia, as I think they provide the best fodder for the non-doxasticist. One might complain that I am cherry-picking my targets, but I prefer to think that I am narrowing down on a natural class of psychological states by noticing what they have in common. I believe that the account I will give explains *most* clinical delusions, and I do not consider it much of a deficiency if it does not capture *all* clinical delusions.

2.1.2 Intensional Characterizations

2.1.2.i Jaspers's Characterization

The first influential intensional characterization of delusion was produced by the phenomenological psychiatrist and existentialist Karl Jaspers in his lauded *General Psychopathology* (1963). The influence of this work on psychiatric theory cannot be overstated.[2]

In early 20th century Germany, Jaspers found himself in an academic milieu that was everywhere engaged with the *Methodenstreit*: a scholarly debate regarding the proper relation between the human sciences (*Geisteswissenschaften*) and the natural sciences (*Naturwissenschaften*). Positivists such as Comte and Durkheim held that the human sciences should strive to emulate the methodology of the more successful natural sciences as much as possible; their opponents, such as Max Weber, thought otherwise. Jaspers was a member of the latter camp; he was, in fact, a member of Weber's inner circle. Weber and Jaspers each thought of their domains of expertise—sociology and psychopathology, respectively—as hybrid disciplines, each of which ought call upon both humanistic and scientific methodologies. This put Jaspers at

[2]For further information about Jaspers on delusion and his influence on psychiatry, see Thornton (2007), Eilan (2000), Garety and Hemsley (1997), and Bentall (2003). Much of the historical information in this section is drawn from these texts.

odds with most of his disciplinary colleagues. Psychological researchers in Germany (such as Wernicke, Alzheimer, and Griesinger) were predominantly interested in exploring the neurobiological underpinnings of mental disorder. Jaspers was in a minority in desiring a more humanistic method for the analysis of psychopathologies. For example, in an early paper, Jaspers considered whether paranoia ought to be considered a biological pathology, or an abnormal but understandable evolution of a patient's personality. In a series of case studies, Jaspers paid particular attention to his patients' subjective reports of the development of their fears. These various papers and studies marked the introduction of the *biographical method* into psychiatry. The biographical method has the psychiatrist attempt to interpret the patient's symptoms as being a part of the patient's life history. Empathy is used as an interpretive tool. Some symptoms can be slotted into a patient's life narrative; these are, in Jaspers's terms, *understandable*. Other symptoms are *ununderstandable*; they can only be explained by referring to an abrupt personality change caused by some biological pathology. This distinction survives in psychiatry today in the distinction between functional and organic pathologies.

The distinction forms the base of Jaspers's theory of delusions. Delusions, he claimed, came in two sorts. Secondary delusions ("delusion-like ideas") are understandable, whereas primary delusions ("delusions proper") are ununderstandable. Jaspers furthermore claimed there were two sorts of understandability: static and genetic understandability. On neither conception are primary delusions understandable. A mental state is statically understandable if the psychiatrist can understand what it would be like to have that state. He must be able to empathetically or imaginatively project himself into the patient's shoes. You can presumably imagine what it would be like to believe that the CIA is hot on your trail, but it's much more difficult to imagine what it would be like to believe that you are dead. The goal of the phenomenologist is to achieve static understanding. Genetic understandability, on the other hand, concerns how one mental state "emerges" from another. We achieve genetic understanding when we can tell a story about how the delusion arose out of the patient's inner psychic life: that is, when we comprehend what other mental

states generated the delusion and are generated by it.

The distinction between primary and secondary delusions has been enormously influential to this day, though it has been less popular in Germany than in Britain and the United States (Winters and Neale (1983), cited in Garety and Hemsley (1997, p.4)), and it has been challenged often. Nonetheless, theories of delusion by modern philosophers often recapitulate something like the distinction. Dominic Murphy, for instance, writes: "Delusions, I suggest, [are attributed] when we run out of the explanatory resources provided to us by our folk understandings of how the mind works" (Murphy 2012, p.22). Davies and Davies (2009) draw a similar distinction between pathologies of belief and pathological beliefs, and Bortolotti and Broome (2008) draw a distinction between authored and un-authored delusions. Bentall (2003, p.28) sees in Jaspers an anticipation of Dennett's distinction between explanations conducted from the intentional stance and explanations conducted from the design stance.

This brings us to Jaspers's second explicit characterization of primary delusions: "The term delusion is *vaguely* applied to all false judgements that share the following external characteristics to a marked, though undefined, degree: *(1)* they are held with an *extraordinary conviction*, with an incomparable, *subjective certainty*; *(2)* there is an *imperviousness* to other experiences and to compelling counter-argument; *(3)* their content is *impossible*" (Jaspers 1963, pp.95–6, italics his).[3]

However, there is an inconsistency between this characterization and the one based on ununderstandability. When discussing genetic understandability, Jaspers writes that because primary delusions cannot be understood, we cannot interpret the words of the delusional individual as expressions of an intentional state. The

[3] Conditions *(1)* and *(2)* are sometimes conflated, though they are distinct. In Bayesian terms, *(1)* corresponds to a degree of belief, and *(2)* corresponds to degree of belief conditional on evidence. These aren't identical: a belief can be very strongly held but be "fragile" to counter-evidence. Andy Egan has suggested to me that knowledge of names is often fragile: I can be very confident what your name is, but if I hear others call you by a different name and you respond, I will quickly change my credences dramatically. (The phrase "strong conviction" can denote either *(1)* or *(2)*, which might lead to confusion.)

apparent assertions of the insane are "senseless ravings" without intentional content (Jaspers 1963, p.577).[4] As Eilan (2000) rightly points out, Jaspers is committed to three inconsistent propositions: (1) Delusions cannot be genetically understood; (2) Delusions have intentional content; (3) A state cannot have intentional content unless it is genetically understandable. A substantial number of debates about schizophrenia, she claims, have boiled down to which of these propositions should be rejected.

Thus, Jaspers's two characterizations of primary delusions—as unshakable convictions on one hand, and as ununderstandable ravings on the other—are at odds with one another. The former characterization treats delusions as beliefs; the latter characterization pushes against thinking of them as such. Even this canonical work from the early days of psychiatry seesaws on whether or not delusions are simply too strange to be beliefs. Non-doxasticism was there at the get-go.

2.1.2.ii Modern Psychiatric Characterizations

In modern psychiatry, the characterization of delusions as unshakeable convictions has enjoyed more popularity. The most obvious place to look for modern characterizations of delusions is in modern psychiatric texts and diagnostic manuals.[5] I offer three examples below.

The first is from the *DSM-IV-TR*:

> *Delusion:* A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not one ordinarily accepted by other members of the person's culture or subculture (e.g. 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 credulity.

---

[4] German Berrios is a contemporary advocate of a "senseless ravings" view of delusions. He claims that delusion reports are "empty speech acts that disguise themselves as beliefs" (Berrios 1991, p.8). I discuss Berrios at further length in a later chapter.

[5] See Garety and Hemsley (1997) for a particularly excellent overview of various characterterizations of delusion that have been offered.

The second is from Mullen (1979, p.39):

> A delusion is an abnormal belief. Delusions arise from disturbed judgments in which the experience of reality becomes a source of new and false meanings. Delusions usually have attributed to them the following characteristics:
> (i) They are held with absolute conviction.
> (ii) They are experienced as self-evident truths usually of great personal significance.
> (iii) They are not amenable to reason or modifiable by experience.
> (iv) Their content is often fantastic or at best inherently unlikely.
> (v) The beliefs are not shared by those of a common or cultural background.

Thirdly, Oltmanns (1988) offers the following features:

> a. The balance of evidence for and against the belief is such that other people find it completely incredible.
> b. The belief is not shared by others.
> c. The belief is held with firm conviction. The person's statements or behaviours are unresponsive to the presentation of evidence contrary to the belief.
> d. The person is preoccupied with (emotionally committed to) the belief and finds it difficult to avoid talking or thinking about it.
> e. The belief involves personal reference, rather than unconventional religious, scientific, or political conviction.
> f. The belief is a source of subjective distress or interfere's with the person's occupational or social functioning.
> g. The person does not report subjective efforts to resist the belief (in contrast to patients with obsessional ideas).

Many writers have noted that these sorts of characterizations fail to precisely capture the nature of delusions.[6] For one, all three of the characterizations above make reference to delusions not being shared by other members of the delusional subject's subculture. This seems suspicious. Many delusions are political or religious in nature, and delusional individuals might well seek out and join extremist groups or conspiracy-oriented internet message boards that roughly comport with their delusional worldviews. It does not seem that one should be able to escape the diagnosis of delusion simply by altering the social environment within which one moves. It is tempting to think that this criterion is a political inclusion meant to paper over any

[6] Given its dominance in psychiatric practice, the *DSM* characterization has received the lion's share of criticism. See Coltheart (2007), Spitzer (1990).

issues that might arise from the implication that common religious or political affiliations are pathological.[7] It would also seem to rule out extreme cases of shared delusions (*folies à plusieurs*) in which entire crowds hold a delusion together, as might happen during mass hysteria.

Secondly, all three mention falsity (as does Jaspers's characterization), yet there is no reason to think that a delusion must be false. If this were so, a therapist could relieve her client of persecutory delusions by secretly following her around. As Jaspers (1963, 109) points out, an individual with the Othello Syndrome, who has an irrational conviction that his or her spouse is unfaithful, could in fact be right. Murphy (2010b) mentions Fulford's (1989) example of an individual who seemed to have only a single monothematic delusion: that he was mentally ill. Amusingly, if delusions were necessarily false, then this would present us with something like a liar paradox.

Thirdly, psychiatric characterizations often make reference to the bizarreness, oddity, impossibility, or fantastical nature of the content. It isn't necessary for delusions to be indisputably bizarre. Again, consider the Othello Syndrome. The proposition "my husband is cheating on me" is hardly bizarre in itself. When spoken by many people, it is true!

However, the mention of these features is not unwarranted. Given the purposes for which these texts are written, the characterizations are far from the lists of necessary and sufficient conditions that philosophers might expect. They are put forth in a *diagnostic* spirit, and can be seen as *symptomatic* of the presence of delusions. Coughs and headaches are not constitutive of the flu, nor are they even present in all flus. But they are reliable indications of a flu. The features listed above can be seen as defeasible diagnostic evidence for the presence of a pathology. If a person is convinced in some proposition that others in the culture see as crazed and fantastic, that gives a hint that something might be wrong, but it certainly doesn't guarantee

[7] See Kirk and Kutchins's *The Selling of DSM: The Rhetoric of Science in Psychiatry* (1992) for a fascinating investigation into the social factors involved in the construction of the third *DSM*. Many of the diagnostic categories were included because of diplomacy, bargaining, and political wheedling rather than critical inquiry.

the person is pathologically deluded. (He could, after all, just be a philosopher.) Psychiatric texts usually say that the features listed are neither necessary nor sufficient. Oltmanns, author of the third characterization above, is explicit about this.

We can distinguish between *operational characterizations* of delusion, which are meant for diagnostic purposes, and *theoretical characterizations*, which are meant to elucidate the nature of delusion for the purposes of theory-building. It is not often clear which of these an author has in mind. Part of the reason is that the distinction between diagnostic symptoms of a pathology and the pathology itself is not always respected in psychiatry. In fact, many think that psychiatric disorders such as depression simply *are* syndromes—collections of observable symptoms—and that there needn't be an underlying pathology that causes all instances of that syndrome. Some disorders should pretty clearly be interpreted in this way. Insomnia, for instance, might be thought to be simply what one has when one has difficulty sleeping when in a normal environment. There needn't be a single type of psychological or biochemical cause that undergirds all token instances of insomnia: one has insomnia if and only if one meets that particular operational characterization. The tendency of psychiatrists to treat *all* psychiatric disorders as syndromes often leads to a conflation between operational and theoretical characterizations.[8] Modern diagnostic manuals treat delusions as beliefs, but this might be because there is pragmatic utility in treating them as such, just as there is pragmatic utility in saying that delusions are not normally shared with one's subculture.

2.2 Evidence for Non-Doxasticism

Given that delusions have been traditionally characterized as beliefs, why have so many been tempted by non-doxasticism? In this section, I describe what I take to be

[8]See Murphy (2010a,b, 2005) for an overview of these issues. The *DSM* is officially agnostic on whether or not the disorder categories it lists are caused by an underlying pathology, and the manual's dominance in the field is part of the reason for there being a lack of interest in underlying pathologies. My own characterization of insomnia was meant to be illustrative. The actual *DSM-IV-R* characterization of insomnia shows an example of symptom-driven characterization: "The *essential* feature of Primary Insomnia is a *complaint* of difficulty initiating or maintaining sleep [...]" (First 2000, 599, my italics). Whatever insomnia is, it isn't essential that one *complain* about not being able to sleep.

the best evidence.

Let us first get clear on exactly what is meant by 'non-doxasticism'. It is usually characterized as the thesis that delusions are not beliefs, but this won't do for a number of reasons. Firstly, it's not necessary to think this in order to count as a non-doxasticist. Some authors who are considered quintessential non-doxasticists do in fact think that delusions are beliefs. For example, Gregory Currie argues that delusions are false metacognitive beliefs. The madman who claims to be an emperor doesn't really believe he's an emperor; he imagines that he's an emperor and mistakes that imagining for a belief. Thus, on a typical reading of Currie, his delusion *is* a belief: a belief that he believes he is an emperor. Rather than suggesting that delusion is an attitude distinct from belief, Currie proposes that delusions have a content other than the one that they *prima facie* appear to have. As Currie is considered one of the primary non-doxasticists, and is motivated by the same sorts of concerns that motivate other non-doxasticists, a characterization of non-doxasticism should include theories that delusions are beliefs with contents other than the ones verbally expressed. This would also bring into the fold those who think that delusional reports are not expressions of a belief with that literal content, but metaphorical expressions of the some other belief.

I propose the following characterization of non-doxasticism. Non-doxasticism about delusions is the conjunction of two theses: a conceptual thesis that that one can have the delusion that p without believing that p, and an empirical thesis that delusional patients in fact do not typically believe the contents of their delusions.[9]

---

[9]There's a class of positions usually considered to be non-doxasticist that this characterization does not happily cover: "senseless ravings" views, such as those of Berrios (1991) and others who claim that delusion reports are empty speech acts that have no content. On these views, it does not make sense to talk about "a delusion that p". We can attempt to bend these contentless accounts to our characterization by saying that on these accounts, 'X has the delusion that p' should be understood as meaning 'X uttered a sentence in a delusional context that has no content but that would be typically be translated by a speaker of X's language as having the meaning p.' This is not perfect, as it doesn't explain what it means for a sentence to be uttered in a "delusional context" such that it has no content, but it's probably incumbent on the holders of such a view to be clearer on these sorts of issues. In any event, I'm not too worried about having to account for these sorts of views, as they seem to go against the data quite badly. Delusional speakers do not spout off utterances at random. Their sentences all seem to be comprehensible and semantically related to one another. Berrios needs to explain why the expressions uttered by delusional subjects apparently bear semantic relations and are apparently all about a particular topic (see Eilan (2000)).

This characterization has an additional benefit. Some of the arguments that have been made in support of non-doxasticism threaten to collapse the debate into a merely terminological one over the proper referent of the word 'delusion'. For instance, some theorists argue that delusions aren't beliefs because there is more to delusion than belief: a delusion comprises a whole collection of phenomenal experiences. This is a merely verbal dispute, and characterizing non-doxasticism as I do above protects us from accidentally taking it to be otherwise.[10]

There are two main sorts of arguments for non-doxasticism: *conceptual arguments* and *explanatory arguments*. According to the conceptual arguments, there are conceptual requirements on belief; a mental state only counts as a belief if it has certain features, but delusions lack those features. According to the explanatory arguments, there are features of delusions that doxasticists cannot easily explain, but a non-doxasticist *competitor theory* can do the trick. Both of these types of argument are motivated by the observation that delusions have certain puzzling features that make them unlike prototypical beliefs. It's to those features that I now turn.

2.2.1 Which Features of Delusion Are Most Puzzling?

Different philosophers focus on different features of delusion when making the case for non-doxasticism. Bayne and Pacherie (2005), for instance, present the following challenges to doxasticism: delusions can have incoherent content; they can be pragmatically self-defeating; they can be founded on insufficient evidence; they can be inconsistent with other beliefs; subjects don't behave as if they believed their delusions; subjects don't exhibit appropriate affective responses. Of course, some delusions exhibit only some of the features mentioned (Bayne and Pacherie note that the Cotard delusion is one of the few that are pragmatically self-defeating).

I would like to call attention to the following four features. For reasons I will present shortly, I believe the latter two are more important than the former two.

[10] I argue for this in chapter 4.

BIZARRENESS OF CONTENT: Delusional content is often bizarre and peculiar in nature, and is usually highly implausible. For instance, it is extremely unlikely that one's spouse has been replaced with an imposter, yet this is what Capgras patients attest.

UNRESPONSIVENESS TO EVIDENCE: Delusions do not respond to evidence in an appropriate or a typical manner. This is a claim about both the formation of delusions and the maintenance of delusions: they are formed in response to insufficient evidence, and they are maintained in spite of evidence to the contrary. For instance, it is often thought that the Capgras delusion is formed in response to a strong experience of unfamiliarity. This is not a rational response. Even if the experience is very powerful, the notion that the patient's spouse has been replaced does not cohere with the patient's other beliefs. The Capgras patient ought reject the proposition, as the evidence for it is insufficient.

CIRCUMSCRIPTION: Delusions are often behaviorally, theoretically, and affectively circumscribed. A typical Capgras patient might continue to live with his spouse, might not call the police, might not form hypotheses about how his spouse was removed and where she has been taken, and might not even be that perturbed about his situation.

DOUBLE-BOOKKEEPING: Delusional subjects are able to keep delusions reasonably well partitioned from beliefs that are not delusions, and are often metacognitively aware of which of their mental states are delusional. Capgras patients will sometimes recognize that their delusions are, in fact, delusions, and will admit that their delusions are incredible and defective in some way. Young reports, "[i]f you ask What would you think if I told you that my wife has been replaced by an impostor?', you will often get answers to the effect that it would be unbelievable, absurd, an indication that you had gone mad" (1998, p.37).

2.2.1.i Bizarreness and Unresponsiveness

Which of the features above provide the best fodder for the non-doxasticist? Which features diverge most sharply from prototypical belief, and demand a novel theory?

Although the bizarreness of delusions is one of their most striking features and is oft-cited in characterizations of delusion, it is at best symptomatic of delusions. In fact, bizarreness is likely not even that diagnostically useful. Researchers have consistently failed to find acceptably high values of inter-rater reliability on judgments of bizarreness (Kendler et al. 1983, Flaum et al. 1991, Junginger et al. 1992). Even if delusions are typically more bizarre than prototypical beliefs for some understanding of bizarreness, they inherit this trait from something more fundamental, and it is this that the non-doxasticist should seek to explain. It's not that delusions are candidates for being non-beliefs because they have contents that most people would find weird; it's that one wouldn't expect the normally reliable mechanisms that form and maintain beliefs to form something that flies in the face of so much other evidence.[11] The problem in explaining the bizarreness of delusions reduces to a problem in explaining their unresponsiveness to evidence.

I don't think that unresponsiveness has much to offer either a conceptual argument or an explanatory argument for non-doxasticism. Let's discuss the conceptual arguments first.[12] Delusions are not formed or maintained as typical beliefs are, and this is supposed to count against delusions counting as beliefs; however, not all causal relations are as important to typing mental states as others. It's the effects of delusions, not the causes of delusions, that are key. Eric Schwitzgebel is a philosopher who agrees that effects of belief, but not the causes of belief, are relevant to its being a belief. He writes, "[b]eliefs can arise in any old weird way, but—if they are

[11] Of course, it's true that normal people can believe some pretty weird things. However, the complete inability of the delusional subject to brook counterargument seems to be of a different order altogether. Some weird convictions of "normal people" probably are held with the same intense degree of irrational certitude as psychotic delusions, but these convictions are in want of explanation just as delusions are. How could a normally reliable epistemic system develop such unyielding irrationality?

[12] In chapter 3, I'll argue that conceptual arguments are unsuccessful. I'm here putting myself in the shoes of a philosopher trying to make the best case for a conceptual argument.

to be beliefs—they cannot have just any old effects. They must have, broadly speaking, belief-like effects; the person in that state must be disposed to act and react, to behave, to feel, and to cognize in the way characteristic of a normal believer-that-P" (Schwitzgebel 2012, p.15).

Schwitzgebel holds this view out of a commitment to thinking of intentional states as dispositions. Consider the formulation of functionalism as first laid forth by Putnam: functional properties are dispositions which take stimuli and other mental states as input, and produce and behavior and other mental states as output.[13] The inputs here are conditions under which the outputs are generated; not conditions that specify when the state is formed. A dispositional profile does not say anything about what it takes to gain that disposition. A cup is fragile if it breaks when struck, and it does not matter what caused it to gain the property that it breaks when struck. On any purely dispositional theory of belief, whether a belief is formed in response to appropriate evidence is not relevant to typing the state.

Of course, dispositionalism is not forced on a philosopher making a conceptual argument. Some versions of functionalism will claim that a state is a belief iff it plays a certain role in a functional economy, and one might hold that the conditions causing a state to arise are important in typing that state. For what it's worth, it does seem to be part of my concept of belief that the belief can arise from anywhere.[14] Suppose a philosopher thinks that conceptual analysis can tell us something about the nature of belief, and she wants to make a conceptual argument for non-doxasticism, she cannot simply assume that the cause of a belief is relevant to the concept of belief; she will need to accommodate my (and Schwitzgebel's) intuition.

Unresponsiveness provides even less fodder for explanatory arguments. Unresponsiveness to evidence is the feature of delusions that has commanded the bulk of attention in philosophy and cognitive science. Over the last thirty years, much of the work on delusions has attempted to answer questions about why delusions are

[13] I am intentionally glossing over the distinction between role functionalism and realizer functionalism (McLaughlin 2006) for reasons of brevity.

[14] Whether the population at large shares this intuition would be an interesting experimental discovery.

formed and why they are not immediately rejected in the face of overwhelming counterevidence. As we will see later, a lot of progress has been made in understanding delusion fixation. These theories have been developed under the assumption that delusions are beliefs, so the fact that delusions are unresponsive to evidence does not provide much motivation for non-doxasticism.

It's the other two features of delusion which I describe below—circumscription and double-bookkeeping—that provide the motivation for explanatory arguments. Although we have several promising theories about what causes delusions to form, we currently have *no* good theories about those two features. What needs explanation is *what delusions do*, not *what gets done to delusions*.[15]

None of this is to say that an account of delusion formation is irrelevant to non-doxasticist theories. It might well be that a non-doxasticist theory will have implications for how delusions are formed. In that case: great! But this will likely be a side benefit to be gained in the search for an explanation of the causes of delusion.

2.2.1.ii Circumscription

Circumscription refers to the relative inertness of delusions in affecting cognition and behavior. Delusions drive their hosts to make bizarre claims, but there are suspiciously few effects beyond the production of verbal behavior. An imprecise slogan: delusional subjects *say* but don't *do*. Three kinds of circumscription can be identified: theoretical, affective, and behavioral.[16] Delusions are theoretically circumscribed in that subjects usually make little attempt to reconcile their non-delusional beliefs with their delusion; they do not seem to incorporate the consequences of

[15]This distinction is often glossed over in discussions of the features of delusion. For example, consider Bortolotti's (2009b) claim that "delusions violate norms of procedural rationality by being badly integrated with the subject's beliefs and other intentional states." There are two features smushed together here: *(1)* unlike beliefs, delusions are not modified to be made consistent with other mental states, and *(2)* other states are not altered to be made consistent with delusions, unlike beliefs.

[16]A note of terminology: 'circumscription' is sometimes used in the literature to refer to the narrower category that I call 'theoretical circumscription'. For example, see (Bortolotti 2009a).

their delusion into their general account of the world. They are affectively circumscribed in that the subject will often not display the expected or the appropriate emotional responses to their delusion. They are behaviorally circumscribed in that delusional subjects will often not act on their delusion in the expected or appropriate way. Could a person really believe that her food is poisoned and yet still continue to eat? Her behavior could be explained if she had other aberrant beliefs and desires, such as a desire to die, or a belief that poison is harmless. Yet patients usually seem to lack these sorts of auxiliary mental states.

The circumscription of delusions is mysterious, but should not be overstated. If you ask a delusional subject why they believe what they do, they might do any number of things. Often, they will spin off an extremely contrived and fantastical answer on the spot. Delusions produce some truly outrageous confabulations, the likes of which are extraordinary.[17] One might think that the fact that delusional subjects confabulate shows that delusions are not as theoretically circumscript as often claimed.

Yet, consider Ramachandran's patient who, when asked where her mother was, said she was hiding under the table. It does not appear that she formed a *belief* about her mother hiding under the table. After all, she didn't behave as if she believed it! If her confabulation did involve the creation of a mental state, it looks like it was a mental state of exactly the same type as her delusion, as it shares all the same problematic features. William Hirstein writes,

> [There is] evidence against the claim that confabulations are expressing real, enduring beliefs [...] confabulators may give different (sometimes inconsistent) responses to the same question at different times. The states that give rise to their claims do not seem to have the permanence of normal beliefs. They are perhaps more momentary whims or fancies than genuine beliefs. As is often the case with the mentally ill, folk psychology is strained to the breaking point as we try to describe them. (Hirstein 2005, p.189)

[17] Confabulations can be either spontaneous or provoked. Only truly serious disorders such as schizophrenia generate spontaneous confabulations.

In fact, it is possible that many delusions are the result of confabulation: the Capgras delusion might be the result of confabulating an explanation for a disordered experience, for instance. A non-doxasticist theory should not simply explain why delusions have little impact on belief revision. It should explain why the mental states that *are* revised by delusions seem themselves to be not quite like normal beliefs.[18]

The problem with explaining circumscription is that delusions are sometimes circumscribed and sometimes not, and it is very hard to predict when or how they will affect behavior or reasoning. For instance, subjects will occasionally feel great consternation over their delusions, and they will also occasionally act on them, sometimes with terrible and tragic results. One man, under the conviction that his father had been replaced with a robot, decapitated him in order to find the batteries and microfilm in his head (Blount 1986, Silva et al. 1989). Behaviors of this sort are thankfully relatively rare. It is also sometimes possible to get delusional subjects to briefly submit to reason; it is difficult to predict how they will react to confounders for their delusions. They might ignore the confounder entirely, they might confabulate a highly improbable justification for their delusion or challenge the confounder, or they might actually correct their mistake, but soon afterward be sucked right back into the delusion (Buchanan et al. 1993). An explanatory theory should attempt to give some sort of account as to when a delusion will prompt action and when it will not.[19]

---

[18]Confabulation is not specific to the mentally ill. Originally, confabulation was thought to necessarily involve memory disorders, but it has become apparent that it's a common, even quotidian, phenomenon. We all confabulate. One famous demonstration is due to Nisbett and Wilson (1977): participants were presented with a series of nylon stockings and asked which they preferred. Owing to a cognitive bias, the majority of participants chose the rightmost pair. When asked why they preferred that one, they made up some explanation, for instance, involving texture or color. However, the results of these confabulations are not in the same boat as the confabulated proposition "my mother is under the table." When regular people confabulate explanations, it looks like the confabulation might generate beliefs: the participants really do think that they chose the stockings because of the texture (Hirstein 2005).

[19]Most non-doxasticist competitor theories tend to simply say that a delusional person will sometimes act on her delusion and sometimes will not. The theoretical benefit of such a claim is fairly minimal. See the upcoming criticism of Egan and Schwitzgebel.

2.2.1.iii Double-Bookkeeping

Double-bookkeeping, also known as *double registration*, refers to a phenomenon first identified by the early psychiatrist Eugen Bleuler (1950) and written about at length by Louis Sass (1994). Delusional subjects appear to keep two "books" or "registries" of representations in their heads—one of "the real world" and one of "the delusional world"—and they are able to keep these books reasonably well partitioned. Surprisingly, delusional individuals often speak as if they are aware that they operate on two different sets of representations: they speak of their "delusional reality" and can identify which of their mental states are delusional. Sass explains,

> it is difficult to square standard notions of poor reality-testing with the fact that many schizophrenics who seem to be profoundly preoccupied with their delusions, and who cannot be swayed from a belief in them, nevertheless treat these same beliefs with what seems a certain distance or irony. [...] It is remarkable to what extent even the most disturbed schizophrenics may retain, even at the height of their psychotic periods, a quite accurate sense of what would generally be considered to be their objective or actual circumstances. Rather than mistaking the imaginary for the real, they often appear to be living in two parallel but separate worlds: consensual reality and the realm of their hallucinations and delusions. (Sass 1994, 20–1)

A person who is deluded is sometimes said to lack "insight" into her disorder. This is sometimes argued for on conceptual grounds: e.g. "being psychotic means being out of touch with real events and experiences. It is therefore incompatible for such people to have knowledge of, or be aware of, true changes taking place within them and the environment" (Marková and Berrios 1992, p.857). Double-bookkeeping challenges the notion that entertaining delusional thoughts implies a lack knowledge of the mundane world.

Psychotic breaks and delusional episodes in people with schizophrenia are often accompanied with a particular kind of feeling—a "delusional atmosphere"—which can aid critically aware sufferers in distinguishing their delusions from their non-delusional beliefs (Fuentenebro and Berrios 1995, Mishara 2010). Sass (1994), Gallagher (2009), and other phenomenologically-inclined philosophers have argued that the experience of the schizophrenic is of living in "two realities." It is difficult to know

quite what this claim amounts to; nonetheless, it is how delusional individuals describe their inner experience. For instance, here is a written report from an individual with schizophrenia:

> I can feel absolutely certain that space and time (and hence physical reality) no longer or never did exist, and yet understand that in order to get to a psychiatry appointment I have to walk down the street, get on the train, and so on (in other words, physically navigate or move through the "objective" world). Or I can feel certain, even as I am talking to my psychiatrist, that I killed him five minutes earlier (fully aware that he is sitting a few feet from me talking). The strangeness is that both "beliefs" exists simultaneously and seem in no way to impinge on one another (nor have I ever figured out any way of consciously reconciling them). (recounted to Louis Sass)

The claim that delusional patients live in "two worlds" is evocative, but must be taken metaphorically. The problem is in figuring out exactly what the metaphor amounts to. Here are three possible interpretations, all of which seem to be supported by the evidence:

BEHAVIOR GUIDANCE: Delusions are behaviorally circumscribed, but this does not mean that patients lie there inert; instead, their actions seem to rely on a *different* knowledge store.

METACOGNITION: Delusional subjects seem to have a *metacognitive knowledge* that there is a division of some sort in what they take to be true.

PHENOMENAL EXPERIENCE: Delusional subjects have a phenomenal experience of "two worlds" overlaid upon one another.[20]

Double-bookkeeping provides some of the best support for an explanatory argument for non-doxasticism. The phenomenon suggests that the delusional mind contains two distinct and partitioned sets of representations, and that the subject appears to have some form of metacognitive awareness of this division. Non-doxasticism

[20]Some reports of double-bookkeeping seem to generate contradictions; perhaps it is possible for conscious experience to contain contradictory contents.

can begin to make sense of this split. An individual might be deluded that *p* but believe that not-*p*, and the individual can know that there is a discrepancy between her delusions and beliefs.

2.3 Summary

Although delusions have traditionally been described as irrationally formed beliefs, certain puzzling features of delusion have long cast doubt on this characterization. Most prominently, delusions are circumscribed—they don't cause the behaviors, inferences, and affective responses that a belief would apparently cause—and they exhibit "double-bookkeeping"—delusional subjects can be aware of their delusions *qua* delusions, and seem to simultaneously experience the world of their delusions and the world as the rest of us know it. These features have led to the rise of non-doxasticism: the thesis that one can have the delusion-that-*p* without believing-that-*p*, and that delusions that *p* are, in fact, often not accompanied with a belief that *p*.

Two types of argument for non-doxasticism have emerged. Conceptual arguments hold that a mental state only counts as a belief if it has certain features, but delusions lack those features. Explanatory arguments hold that the puzzling features of delusion can be better explained by a competitor theory that does not treat delusions as beliefs.

In the next chapter, I argue that too much attention has been paid to conceptual arguments by doxasticists and non-doxasticists alike. I do not see much hope for a conceptual non-doxasticist argument; non-doxasticists must try to build competitor theories.

Chapter 3
Conceptual Arguments and Rationality

3.1 The Irrationality Argument

The features of delusion presented in the previous chapter have caused philosophers and psychologists to be skeptical that delusional utterances are expressions of belief. Delusions aren't like prototypical beliefs, certainly—but are they different *enough* for the non-doxasticist to be right? Say that you describe a person's mental state by describing the way it behaves and the way it interacts with other mental states—that is, by describing its functional role. What determines whether that state is a belief? How do you *type* mental states?[1] It would be handy to have a list of criteria that specify the functional role of belief. Then, if delusions lacked a crucial feature, we would know that they are not beliefs.

Arguments proposing that delusions lack one or more of these necessary features are *conceptual arguments* for non-doxasticism. Most non-doxasticist arguments are conceptual arguments. They have the following form: *Beliefs necessarily have feature C. Delusions lack C. So, delusions are not beliefs.* I call these "*conceptual*" arguments because the first premise is often justified through a priori conceptual analysis: analysis of BELIEF tells us something about the necessary features of belief. In fact, the features that are elicited through this process are often claimed to be not just necessary features of belief, but features that are *constitutive* of belief. In the metaphysician's parlance, they are features *in virtue of which* a state is a belief, or features that *ground* a state's beliefhood.

Determining the constitutive features of belief has kept philosophers occupied

[1] I use the verb '*to type*' as shorthand for '*to determine the type of*'.

for a long time, and there are a lot of proposals on the table. For example, some claim that it is constitutive of beliefs that they cannot be formed voluntarily (Williams 1973). Others claim that it is constitutive of beliefs that they be stable and resistant to reconsideration (Holton forthcoming). Others present normative features, and claim that it is constitutive of belief that it have truth as its aim (Wedgwood 2002, Chan forthcoming).

One path to constructing a conceptual argument is to pick out one or more allegedly constitutive features of belief and plug them into *C* in the argument schema above. For example, Berrios gives a conceptual argument when he writes,

> We must now test the hypothesis that, from the structural point of view, delusions are, in fact, beliefs. Price (1934) distinguished four elements that comprise a belief (P):
>
> a) Entertaining P, together with one or more alternative propositions Q and R;
> b) Knowing a fact or set of facts (F), which is relevant to P, Q and R;
> c) Knowing the F makes P more likely than Q or R;
> d) Assenting to P; which in turn includes (i) the preferring of P to Q or R; (ii) the feeling of a certain degree of confidence with regard to P.
>
> Price's criteria are clear and elegant enough, but it is clear that no current textbook or empirical definition of delusion can be set in terms of these four criteria. (Berrios 1991, p.8)

The most prominent conceptual argument for non-doxasticism appeals to rationality. In virtue of being unresponsive to evidence and in virtue of being circumscribed, delusions are not rational. (It is less obvious that double-bookkeeping is irrational, partly because it is so hard to explain exactly what is going on in double-bookkeeping.) If beliefs are necessarily or constitutively rational but delusions are irrational, then delusions are not beliefs. They are simply *too irrational* to count as beliefs.[2]

[2] One might quibble here: a non-doxasticist would not be able to talk about "the irrationality of delusions," as irrationality is a feature of beliefs, sets of beliefs, or belief-forming processes, and the non-doxasticist denies that delusions are beliefs. A non-doxasticist could also not say things such as 'if one were to interpret delusions as beliefs, they would be massively irrational beliefs', for the non-doxasticist who believes in rationality requirements denies that there can be massively irrational beliefs. I'll continue to use locutions such as these throughout this thesis, as they are a natural way of speaking and do not lead to any important confusions that I can see. The irrationality of a delusion should be understood as follows: a delusion is irrational if it is not a rational belief.

Let's call this particular conceptual argument *the irrationality argument*. I'll eventually speak against conceptual arguments in general, but first, I plan to take down this one, the most prominent. By doing so, we will get a sense of how conceptual arguments in general can be resisted. In this chapter, I argue against two sorts of non-doxasticist argument that posit a conceptual link between belief and rationality, which will lead to some general suspicions about conceptual arguments in general.

I begin by introducing the debate over the irrationality argument by considering concerns that have already been voiced by the doxasticist opponent.

3.2 Delusions as Irrational Beliefs

3.2.1 Bortolotti on Irrationality

Appeals to rationality constraints on belief are plentiful in the non-doxasticist literature. Currie and Ravenscroft, for example, maintain that beliefs must exhibit what they take to be one element of rationality: the consistency of beliefs is "a minimal requirement on belief of any kind" (2002, p.176). They use this as a premise in an argument against doxasticism.[3] There is history and pedigree in the claim that beliefs are necessarily rational. Arguments and thought experiments that have been mustered in the past can now be conscripted to the non-doxasticist's side. Consider, for example, a thought experiment of Daniel Dennett's (1987a, p.44). Dennett describes a "neurocryptographer" attempting to insert a new belief into the brain of a test subject named Tom: 'I have an older brother living in Cleveland'. Either the belief will be rejected for conflicting with other beliefs, or Tom's rationality will be so impaired that we can't take him to be a genuine believer: "in neither case has our neurocryptographer succeeded in writing a new belief." As Coltheart and Davies note, Dennett's vivid thought experiment is "clearly relevant to the topic of delusional beliefs" (Davies and Coltheart 2000, p.3).

Because there is a past literature to draw upon, non-doxasticists do not often argue for rationality as constitutive of belief. Instead, they cite previous theorists who

[3] See Berrios (1991), Campbell (2001), Stephens and Graham (2004) for other similar arguments.

have made the claim. Dennett, Davidson, and Lewis are all popular figures to invoke.

We find more explicit arguments by looking across the aisle. In response to the rising tide of non-doxasticism, certain philosophers have begun to defend the standard doxastic account of delusions (Bayne and Pacherie 2005, Bortolotti 2009b). The most thoroughgoing and sustained defense of doxasticism is Bortolotti's recent and important book *Delusions and Other Irrational Beliefs* (2009b). The book is an influential doxasticist missive aimed at the irrationality argument; it is admirably clear, and it nicely summarizes and encapsulates the state of play.

Bortolotti argues that rationality cannot be constitutive of belief, for irrational beliefs are not just possible; they are prevalent. There is nothing incoherent about an irrational belief, so delusions can be considered irrational beliefs, and their odd features can be explained in terms of irrationality. Self-proclaimed kings and emperors in a psychiatric ward might not expect tribute befitting their station, but this should not be taken to straightforwardly imply that they don't actually think they are royalty. Rather, it's natural to say that they do believe that they are monarchs, but they irrationally do not recognize the absurdity of their situation. Circumscription and unresponsiveness might distinguish delusions from *prototypically rational* beliefs, but not from irrational beliefs that are irregular yet mundane.

Bortolotti identifies five types of conceptual argument for non-doxasticism, each of which takes as a premise that adherence to some sort of norm of rationality is a necessary component of belief:

> *Bad Integration.* If delusions violate norms of procedural rationality, by being badly integrated with the subject's beliefs and other intentional states, then they are not beliefs.
>
> *Lack of Support.* If delusions violate norms of epistemic rationality, by being formed on the basis of insufficient evidence, then they are not beliefs.
>
> *Unresponsiveness to Evidence.* If delusions violate norms of epistemic rationality, by resisting revision in the face of counter-evidence, then they are not beliefs.
>
> *Failure of Action Guidance.* If subjects with delusions violate norms of agential rationality, by failing to act on the content of their delusions in the relevant circumstances, then they are not ascribed beliefs.
>
> *Failure of Reason Giving.* If subjects with delusions violate norms of agential rationality, by not endorsing the content of the delusion on the basis of good

reasons, then they are not ascribed beliefs. (Bortolotti 2009b, p.56)

Bortolotti then sets out to attack all of these conditionals, making much of findings in psychology and behavioral economics demonstrating various ways in which we are all irrational. For instance, she calls upon experiments of Kahneman and Tversky (1982) that demonstrate the natural human tendency to make various errors in probabilistic or conditional reasoning. This has been a traditional response to any strong emphasis on rationality constraints on belief ascription: without some fancy footwork, the claim that individuals with beliefs are necessarily rational simply looks empirically false on the face of it (Stich 1985, Thagard and Nisbett 1983, Cherniak 1986).[4]

Especially in the introductory sections of her book, it's clear that Bortolotti takes an attack on rationality requirements of belief to be one of the more important tasks in her book. She writes, "[m]y project here can be seen as an instance of the following strategy. Take a well-established principle. Instead of accepting it as an a priori truth and applying it to the case at hand, start thinking about possible counterexamples [...and ...] start entertaining the possibility that the principle should be revised, replaced by a more refined principle, or altogether abandoned" (Bortolotti 2009b, p.8). Although she denies that beliefs are ideally rational, she finds it at least a little compelling that there is a weaker condition. Defenders of doxasticism tend to agree that there is something *like* a rationality constraint on mental states. Bortolotti endorses what she calls an *intelligibility requirement*: "intentional behavior must be intelligible or amenable to rationalization" (Bortolotti 2009b, p.100). This requirement, she maintains, is weak enough that delusions will count as beliefs.

In order to attack each of the conditionals mentioned above, Bortolotti employs an argumentative strategy that I will call her *central strategy*. For any purported failure of rationality that delusions exhibit, Bortolotti attempts to offer an example of a clear-cut belief that has the same relevant features. For example, she invites the reader to compare the following two cases:

---

[4]Of course, some people have tried to engage in this sort of fancy footwork. See Cohen (1981) for a classic attempt, but see Stich (1985) for a response.

- Whenever Nishad takes an exam, he wears the chain that his grandmother gave him, because he believes that the chain brings him luck and protects him from harm. He knows that in general, objects have no special power on people or situations, but he makes an exception in this case.
- Bob believes that the person who lives in his house and looks identical to his mum is not his mum, but a cleverly disguised Martian, and develops a growing sense of hostility towards the imposter. (Bortolotti 2009b, p.93)

The former is an normal superstition and the latter is a clinical delusion, yet in many respects the two look extremely similar. There does not seem to be an easy-to-state feature *C* that distinguishes the two of them. This gives us reason to think they are instances of the same mental state type. Bortolotti supports a version of the "continuity thesis" (Bortolotti 2011b, Bentall 2003): there is no real categorical difference to be drawn between delusions and mundane superstitions and ideological beliefs. The *DSM-IV-TR* had to include an escape clause in its definition of delusion in order to exclude religious conviction and other socially prevalent belief: all well and good for a clinical manual, but the non-doxasticist offering a theory of cognition cannot get away with *ad hoc* escape clauses. Beliefs can be more or less rational, on Bortolotti's picture, and delusions exist far down on the irrationality end of the scale, nestled alongside beliefs in alien abduction and ghosts and the healing power of crystals.

3.2.2 A Non-Doxasticist Response

Bortolotti's central strategy is powerful: it provides a schema for producing arguments against *any* variety of non-doxasticism, not just those founded on the irrationality argument. The strategy challenges the non-doxasticist to identify a feature that cleanly cleaves delusion from belief—one that shows why Nishad, but not Bob, counts as a true believer. Either Bob's delusion is a belief, or Nishad doesn't really believe the chain will bring him luck, or the two cases are not relevantly similar. It does not look easy to find a relevant property that distinguishes their mental states, whether or not that property has anything to do with rationality.

Perhaps it is possible to articulate such a property, but I share Bortolotti's skepticism. Nonetheless, I do not think Bortolotti's central strategy will much faze a committed non-doxasticist. Given that the non-doxasticist is committed to Bob's delusion not being a belief, and the cases certainly seem relevantly similar in many ways, her best option is to *deny that Nishad believes the chain will bring him luck*. The central strategy hinges on our accepting that the "indisputably" irrational beliefs that are offered really are genuinely irrational beliefs. This can be denied. A non-doxasticist might claim that many mental states in a normal population that *look like* irrational beliefs are, truthfully, *not* irrational beliefs, either by denying that they really are irrational, or by denying that they really are beliefs.

It's not surprising that someone invoking the irrationality argument would make this claim. After all, a philosopher claiming that beliefs are necessarily rational will not grant that superstitions and the like are irrational beliefs. To assume they'd accept this is to beg the question.

I'll say more about the viability of this strategy—claiming that many things that we normally consider mundane beliefs are actually not beliefs—at the end of this chapter. What we need, in order to directly engage the non-doxasticist, is something that Bortolotti's central strategy tried to circumvent: a head-on debate about whether beliefs are necessarily rational. Occasionally, one gets the sense that some philosophers think the very idea of belief requiring rationality is not just wrong; they think it is absurd. Because of this, the force of their arguments often comes down to their ability to convey goggle-eyed incredulity through text. As much as I think that it is sometimes appropriate to give an incredulous stare in place of an argument, the irrationality argument is widespread and I fear that no stare will be piercing enough.

The central strategy does not consider *why* rationality has been said to be constitutive of belief, so Botolotti does not give *counter*arguments to arguments in favor of constitutive rationality. This isn't surprising, given that non-doxasticists tend not to simply refer to a past philosophical tradition, but it has left the irrationality argument relatively untouched. In what follows, I will give the counterarguments that I think are needed. Non-doxasticism should not be founded upon conceptual arguments

that depend rationality requirements, so my goal is to give a hand to Bortolotti and her fellow doxasticists by arguing against those versions of non-doxasticism.

The rest of the chapter proceeds as follows. Firstly, I consider arguments that rationality is a requirement on belief ascription. I consider two sorts of ascription: individual ascription and scientific ascription. In both cases, I find the need for a rationality requirement lacking. Secondly, I consider arguments that rationality is a constitutive feature of the functional role of belief. Again, I argue that rationality is not a requisite feature.

3.3 Rationality Constraints on Belief Ascription

There are two ways to get to the claim that beliefs are constitutively rational. They are not always distinguished.

Firstly, one might hold that the process of interpretation involved in *mental state ascription* is governed by a principle that guarantees the rationality of the interpreted agent. In order to ascribe mental states to another person, an assumption of rationality must be made. Alternately, one might hold that it is characteristic of the *functional role* of belief that it is rational: if a mental state doesn't play the role of a rationally formed and maintained belief that motivates behavior in a rational way, then it doesn't play the role of a belief. On the first thesis, rationality is a condition on an interpretive process. There is an *assumption* of rationality built into the very game of ascribing beliefs. On the second, rationality is a feature of the *outputs* of the process of interpretation: the states themselves.

When non-doxasticists such as Currie and Campbell laud figures such as Dennett and Davidson, they are calling upon arguments of the first sort—those about ascription. They do not, however, flag their commitments as such.

Let's take these arguments in turn. If there are constraints on mental state ascription that guarantee our beliefs will be rational, then one cannot claim that delusions are highly irrational beliefs. In the rest of this section, I consider whether belief ascription demands an assumption of rationality that is sufficiently strong to make a

case for non-doxasticism. Defenders of doxasticism such as Bortolotti, Bayne, and Pacherie claim that there is not (Bayne and Pacherie 2004). I concur, but for reasons different than those that they offer.

3.3.1 A History of Rationality Constraints

Rationality constraints on belief ascription feature most predominantly in a tradition stemming from Quine and his notion of linguistic interpretation. Quine famously argued that translation between languages would always be beset with indeterminacy, but he also claimed that we ought abide by certain maxims of translation that would have us prefer certain translation manuals to others. One such maxim is a "principle of charity" that would have us rule out translations resulting in logical silliness. Take Quine's field linguist, charged with translating a language he has never heard before. He notices that speakers always assent to utterances of the form $\ulcorner q$ ka bu $q\urcorner$. This counts as evidence against translating 'ka' as 'and' and 'bu' as 'not'. The principle of charity is what motivates Quine's (1960) famous declaration that "prelogicality is a myth of bad translators."

Those following in Quine's footsteps, including Davidson, Dennett, and Lewis, took the principle of charity to be associated with projects wider than just linguistic translation. For Davidson and Lewis, the principle of charity is a constraint that preserves rationality during radical interpretation. Radical interpretation is unlike radical translation in that it also ascribes mental states to an agent and is not purely linguistic. For Dennett, rationality of the interpreted agent is an assumption that underlies "application of the intentional stance" (Dennett's preferred way of talking about mental state ascription). For example, Dennett writes, in an argument that is characteristic of this Quinean lineage,

> The assumption that something is an intentional system is the assumption that it is rational; that is, one gets nowhere with the assumption that entity x has beliefs p,q,r... unless one also supposes that x believes what follows from p,q,r,...; otherwise, there is no way of ruling out the prediction that x will, in the face of its

> beliefs p,q,r... do something utterly stupid, and, if we cannot *rule out* that prediction, we will have acquired no predictive power at all.[5] (Dennett 1978a, p.17, my italics)

Just as Quine introduced the principle to rule out translation that would impute logical silliness, the reason for introducing a rationality constraint is to pare down on an otherwise unbridled indeterminacy that would plague mental state ascription. For instance, suppose we are to ascribe mental states to an individual taking an umbrella as she leaves the house. We could attribute to her a belief that it will rain, a belief that taking an umbrella will keep rain from falling on her, and a desire to stay dry. Or, we could attribute a belief that it absolutely will not rain, a belief that an umbrella keeps rain off her, and a desire to stay dry, thus interpreting her as being prudentially irrational.[6]

---

[5] A couple of similar quotes from the Quinean lineage: "When we are not [rational], the cases defy description in terms of ordinary belief and desire" (Dennett 1987b, p.87). "If we are intelligibly to attribute attitudes and beliefs ... then we are committed to finding, in the pattern of behavior, belief, and desire, a large degree of rationality and consistency" (Davidson 1982a, p.237).

[6] It is worth nothing that these Quinean philosophers maintain that the *content* of our mental states is determined through a process of interpretation. Call this *interpretivism*. Many non-doxasticists who make the irrationality argument bind themselves tightly to interpretivism (e.g. see Campbell (2001) and my comments on him in a later chapter). However, this is a stronger claim than is needed in order to make the Quinean argument for a rationality assumption. Non-interpretivists (such as Fodor or Dretske) do not think that content is fixed by interpretation, but the arguments in favor of a rationality requirement are still directed at them. A theory of content on its own has nothing to say about the possibility of irrational organisms: the theory must also be fitted with a theory of *attitude* ascription. For a functionalist, this theory will involve interpretation, even if her theory of content does not involve interpretation. So, for example, one could agree with Fodor that the content of our thoughts is fixed through asymmetric dependence. However, the typing of the *attitude* with that particular content will require a different sort of theory: very possibly one that depends on a process of interpretation. Suppose that a computer or a brain contains an inscription written in Mentalese. Is this particular Mentalese sentence in a "belief box"? Or is it in an "imagination box" and the agent irrationally acts as if her imaginations are beliefs? Non-interpretivists find themselves in the same boat as interpretivists: if interpretivists need a principle to pare away attitude ascriptions, so too do non-interpretivists.

Note that this reading of the non-interpretivist's situation does not rule against the possibility of a punctate *language* (i.e. there could be a language with but a single term or expression), but strongly suggests against the possibility of a punctate *mind* (i.e. there could not be a creature or system with but a single belief). Why should we interpret the agent with a single Mentalese token in its head as *believing* the content of that token? Fodor agrees that "the holism of belief doesn't entail the holism of content" (Fodor and Lepore 1992, 128). It's not clear to me why Fodor doesn't make this move in response to Stich's Mrs. T thought experiment (1996), in which a woman assents to the claim that McKinley was assassinated while also being unable to say anything else related to assassination. Fodor should, I believe, say that although Mrs. T has the concept ASSASSINATED (fixed by asymmetric dependence), it languishes in her head without playing a role in any of her *beliefs*. This is not Fodor's response (see Fodor (1987, p.62)), but it is consistent with it. In any event, this shows that the issues that motivate rationality assumptions are not the concern of solely interpretivists: any functionalist will need to come to terms with them.

In addition to the Quinean tradition, a second major source of historical support for rationality requirements came from formal models of economic behavior. Standard decision theoretic or game theoretic models, such as those put forth by Von Neumann and Morgenstern in *The Theory of Games and Economic Behavior* (1944), are descriptive of human behavior only if in accord with the dictates of the theory, which are usually norms of rationality. Davidson, for one, was heavily influenced by Ramsey's "Truth and Probability" (Ramsey 1931). Ramsey gives a procedure for representing an agent's utilities and degree of beliefs in any proposition when simply given that agent's preferences; he then gives a representation theorem proving that if the agent's preferences satisfy certain requirements, the agent's degrees of belief will be coherent. Davidson took radical interpretation to involve something like Ramsey's procedure.[7] Dennett did too: taking up the intentional stance involves interpreting an agent to have beliefs and desires "roughly as Bayes would have them" (Dennett 1978a, p.307). The constraints of rationality touted by Davidson and Dennett result from demanding that mental state ascription involve a procedure akin to Ramsey's. Interpreting individuals through a formal theory that guarantees the ascription of rational beliefs is much the same as adopting a rationality constraint.[8]

3.3.2 Do We Need a Rationality Constraint?

The above sorts of arguments are illustrative. They make it *seem* like we need rationality constraints to get interpretation off the ground. But they are not decisive, and to my mind, the arguments fail to satisfyingly answer the following two questions:

---

[7] Davidson saw various affinities between Ramsey's procedure and Quinean radical translation. For example, he writes, "Quine's solution resembles Ramsey's, in principle if not in detail. The crucial step in both cases is to find a way to hold one factor steady in certain situations while determining the other" (Davidson 1990, p.319). See Rawling (2003) for more on Quine and Ramsey's influence on Davidsonian radical interpretation.

[8] I'm glossing over questions about what it means to be rational. Is a subject who forms beliefs for pragmatic but non-epistemic reasons considered rational? Does the constraint guarantee both theoretical rationality in one's beliefs and practical rationality in one's actions, or just one or the other? Is it irrational to strongly desire something that you know will never come to be? Can desires be rational and irrational, or only beliefs? Many different versions of the principle of charity have been offered, all of which differ on questions such as these. I shrug off details here because I'm not proposing a rationality constraint of my own, but a supporter of a rationality constraint will need to give answers.

1. Must a *rationality* constraint be one of the constraints on psychological interpretation?

2. Is a constraint on psychological interpretation really required *at all*?

The line of ancestry running through Ramsey provides some apparent reason to think that, if there is a constraint, it is a rationality constraint. The line of ancestry running through Quine provides some apparent reason to think that some sort of constraint is needed, but it doesn't provide much reason to think that the constraint needs to specifically be about rationality.

Let's deal with the Ramseyan line first. The support offered by the Ramseyan line is merely apparent. Humans can be formally represented as rational, but this means nothing in itself, for they can also be formally represented as irrational. As Lyle Zynda (2000) has proved, for any preference ordering that allows one to be representable as having degrees of belief that obey the laws of probability, that same preference ordering allows one to be representable as having of degrees of belief that *don't* conform to the laws of probability. In order to establish that humans are rational, it is not enough to simply establish that humans are representable as having consistent and rational beliefs; there are other representations that say otherwise. Why should we think that the best formal theory of mental state ascription should be contained within the set of formal theories that guarantee rational beliefs? There are other formal models: some posit mental states other than belief and desire (such as intention or emotion); some do not assume that our preference ordering is transitive; some allow for unsharp probability functions. Perhaps a model that does not guarantee rationality will do a better explanatory job. Since the original suggestion that unelaborated Ramseyan decision theory could be used as an empirical model of actual human decision-making (Edwards 1954), the claim has been steadily attacked; psychologists and behavioral economists have developed competing accounts of decision-making and competing research programs. If an interpretive principle that limits indeterminacy is needed, we needn't simply take on the most simplistic decision theoretic models at hand.

Ramsey's representation theorem may have been a motivating influence on the widespread acceptance of a rationality constraint, but its sheer existence cannot be pointed to for the justification of a constraint.

It's worth subjecting Dennett's intentional stance to this sort of criticism. Taking up the intentional stance involves interpreting an agent as having coherent and rational degrees of belief. Now, people obviously don't act exactly like perfect Bayesian agents all the time, and Dennett admits this, but he maintains that this doesn't imply the surprising fact that no one is a believer. The intentional stance is still useful in predicting behavior, and this is because people closely *resemble* Bayesian agents in important respects (Dennett 1991). The property *being a believer* is somewhat like the property *being a rabbit-shaped image*. Some images only vaguely resemble rabbits; others might be smudgy or pixellated. As the fidelity of the image goes down and noise is introduced, it becomes less of a perfect rabbit image, but it still has the same basic pattern that a perfect image would. People are, metaphorically, "smudgy images" of fully rational Bayesian agents. To ask whether a schizophrenic *really believes* that there is a nuclear reactor in his head is akin to asking whether a shape in a smudgy picture *really is* rabbit-shaped. It's like a rabbit image in some respects but not in others—its status is indeterminate and there is no fact of the matter.[9]

Let's grant this story. Taking up the intentional stance is just representing or modeling an individual as having coherent degrees of belief. Now, however, recall that there are other cognitive models and representations waiting in the wings. Consider a new stance—a "schmintentional stance"—according to which individuals are represented or modeled by some different formal structure. Perhaps this representation does not preserve rationality, or perhaps it posits mental states that the intentional stance does not. These models might very well do a better job of predicting human behavior. Dennett often speaks as if, when an individual can't profitably be understood on the intentional stance, we need to plunge down to the design stance or

[9]The metaphysics here are difficult. Whether this account demands ontic vagueness is an open question; accounts of indeterminacy that are purely linguistic don't seem to capture what Dennett has in mind. I'll leave issues related to Dennett's "metaphysics of real patterns" to the side. (Because of this, my language will be rather metaphorical; hence the smudgy talk.)

physical stance. But why? Why not look for models of human psychology that are similar to (but distinct from) the one applied by the intentional stance, but also not steeped in the teleology of the design stance?

We should not conclude that humans must be interpreted according to some psychological model just because they *can* be successfully interpreted according to that psychological model. There might be a better model out there. So, why should we assume the best model guarantees that agents are perfectly rational?[10]

Perfect rationality is an awfully strong demand. Many fans of rationality constraints have agreed, and instead offered constraints that demand less than complete consistency and closure in one's beliefs. Lewis (1974) and Papineau (1991) both subscribe to versions of Grandy's (1973) "Principle of Humanity," which states, roughly: interpret agents to have the sorts of beliefs that humans would have.

As previously mentioned, Bortolotti proposes an *intelligibility requirement*: "intentional behavior must be intelligible or amenable to rationalization" (Bortolotti 2009b, p.100). She suggests that we should consider the interpreter's assumptions about intelligibility to be "flexible and revisable heuristics, not constraints. They are supposed to guide the interpreter and help her to ascribe intentional states with determinate content to a variety of subjects in a variety of situations" (Bortolotti 2009b, p.107). Similarly, Bayne and Pacherie (2004) praise Cherniak's (1986) *minimal rationality constraint* in support of doxasticism. This is a weakened demand that states that, if A has a particular belief-desire set, A would undertake some, but not necessarily all, of those actions that are apparently appropriate. The minimal rationality requirement can, allegedly, accommodate the fact that human beings are in "the *finitary predicament* of having fixed limits on their cognitive capacities and the time available to them" (Cherniak 1986, p.8).

---

[10] Dennett has another argument for a rationality constraint: natural selection, he claims, would prefer rational agents, as rationality contributes to fitness, so we should expect people to be inherently rational. Firstly, it is not clear that rationality really is fitness-inducing (Stich 1985). Secondly, even if we should expect people to be rational because they are evolved entities, this is quite different than saying that rationality is a *condition* on interpretation. I develop this argument in section 3.3.3.ii.

Cherniak's minimal rationality constraint and Bortolotti's intelligibility requirement are not adequate alternatives to the sort of rationality constraint that non-doxasticists appeal to. To see why, let's consider the second of the questions with which I opened this section: do we require any sort of constraint on psychological interpretation at all?

3.3.3 Do We Need Any Sort of Constraint?

3.3.3.i Individual Ascription

There are two tasks one might be undertaking when one attempts to explain mental state ascription. Firstly, one might wish to give a psychological theory explaining how individual human agents ascribe mental states to other human agents. This is often called 'mindreading'. Secondly, one might wish to explain how psychologists and scientists develop theories that ascribe mental states. Let's call the first type of ascription *individual ascription*, and the second type *scientific ascription*. The first is employed by individuals in real-world situations; the second is employed by the psychological community in scientific theorizing. An investigation into either sort of ascription can have a descriptive or normative focus. One might be interested in how individuals actually do go about mindreading, or one can make suggestions about how people ought to mindread. Similarly, one can describe how psychologists build theories that attribute mental states to observed actors, or one can offer suggestions about how their theory-building ought to go. Investigations into individual ascription are traditionally descriptive; investigations into scientific ascription are traditionally normative.

These two projects are distinct. They are also often conflated. As Goldman notes (2006), Dennett and Davidson seem to take a theory of mindreading to be identical with a theory of the metaphysics of mental states. Their commitments to the metaphysics of mental states leads them to reject simulationism (a theory of *individual* mental state ascription) right off the bat. Although Dennett and Davidson may think

that individual acts of mental state ascription involve the application of a tacit theory, the bulk of their writings are more thoroughly engaged with the metaphysical project, and most of their arguments are in support of it.

The project of scientific ascription must be the one that non-doxasticists are engaging with. They are interested in finding out what mental states delusional subjects actually have. They aren't interested in finding out whether individuals ascribe mental states to delusional subjects through application of a folk psychological theory or otherwise. This should be clear from the non-doxasticists' positive stories: many of them claim that delusions are newly-introduced mental states, such as "bimagination," that are not part of our folk psychology. The folk *do* say that delusions are beliefs (Rose et al. manuscript); this shouldn't in itself be taken as a refutation of non-doxasticism.

In arguing against a strong rationality requirement, Bortolotti assumes that the non-doxasticist is engaged in making a proposal about the first sort of ascription—individual ascription—and that rationality constraints are a feature of mental state ascription as performed by actual humans. For example, she writes, "One way of describing interpretation is to say that when an intentional state is ascribed to someone, it is ascribed in virtue of casual relations between an *interpreter* (the individual who does the ascribing), the *subject* of interpretation (the individual whose behavior is being observed), and the *environment* shared by the two" (Bortolotti 2009b, p.9). Or: "My suggestion is that we should view the practice of interpretation as guided by fallible (but generally successful) heuristics" (Bortolotti 2009b, p.106). Many of her comments decrying a strong rationality requirement depend on the observation that *we* do not seem to demand perfect rationality when ascribing mental states. She posits "intelligibility" requirements rather than the stronger rationality requirements in order to explain actual ascription.

This does not engage with the non-doxasticist's project.

3.3.3.ii Scientific Ascription

Let's put theories of mindreading to one side and consider scientific ascription. This is a sort of mental state ascription that the non-doxasticist should be interested in, for the non-doxasticist wants to make a claim about what science ought to posit: our best psychological theories or models should not present delusions as beliefs. Thus, a non-doxasticist might claim that *psychological theory construction* is constrained by a rationality requirement on belief; the rationality constraint is needed to resolve indeterminacy that would otherwise plague theory construction. Is there any reason to think that rationality constraints (or anything like rationality constraints) are required by our best *scientific* theory of the mind?

Consider the observational data we acquire when building theories of physics. We take measurements, we construct atom chambers and run experiments, we build instruments, etc. The actual theory we construct is underdetermined by all this data. We posit atoms and subatomic particles, but an evil demon manipulating all our observations will fit the data equally well. What prevents us from inviting in rampant indeterminacy in our commitments are certain epistemic principles or scientific virtues that guide our theorizing: simplicity, conservatism, scope, fecundity, and so on. If rationality were a constraint on mental state ascription, it would be serving as another such scientific virtue. It would be another such principle that we would use to reduce indeterminacy.[11]

However, there is something very odd about a rationality principle. It constrains only a particular special science: psychology (and perhaps economics). No other sciences seem to require an additional virtue. This should make us suspicious of its necessity. In fact, it's not clear that the *other* virtues cannot do the indeterminacy-reducing job that we want a rationality requirement to do.

I maintain that it is often predictive and profitable to attribute rationality (rather than irrationality) to an agent without appealing to a rationality requirement. This is because ascribing irrationality to a subject is wholly uninformative. To interpret

[11] The virtues listed above are some of those presented by Quine and Ullian (1970).

a person crossing the street as irrational and as *not* wanting to do so doesn't predict much else about them. Why will they say they crossed the street? Would they cross if they wanted to? The theory doesn't say. Attributing rationality and the desire to cross the street, on the other hand, offers up wealth of other information about their potential behaviors in various situations. Thus, we have reason to avoid attributions of irrationality, and the reasons given *do not appeal to a rationality constraint*.

Consider a version of Quine's field linguist, who, seeing speakers assent to ⌜p ka q⌝ when they assent to some p and dissent from some q, prefers to translate 'ka' as 'or' rather than as 'and'. Why should he prefer this hypothesis? On one account, it would be because translating it as 'and' violates a requirement of rationality. Can we get the same result without appealing to such a constraint? If we posit that 'ka' means 'or' and that the speaker is rational, we end up making all sorts of other predictions. For one, we anticipate that he will accept ⌜r ka s⌝ if he accepts any r or s. The hypothesis systematizes a whole lot of possible data about the speaker's dispositions. On the other hand, if we posit that 'ka' means 'and' and that the speaker is irrational, and if we don't have a theory about how the speaker is irrational, then we can't predict much else. We don't know how the speaker will respond to pretty much any instance of ⌜r ka s.⌝ Thus, whatever scientific virtues push one to prefer simple and predictive systematizations of the facts will suggest a theory in which the agent is rational. We have a theory that tells us what can be expected when an agent is rational; simply claiming that an agent is irrational jettisons all those predictions. If a patient has the delusion that he is Napoleon, we can predict at least *some* things about his behavior (such as the fact that he will say that he is Napoleon). If we simply say that the agent has the irrational belief that he is Napoleon, then we should be hesitant to draw any conclusions whatsoever. We lose information. It's the epistemic virtues of predictiveness and systematization that should make us wary of attributing irrational beliefs, and *not* a distinct rationality requirement.

However, if we have a *theory* of agential irrationality—one which gives us specific predictions about how an agent will irrationally act in various scenarios—then these

same virtues can prefer attributions of irrationality to attributions of rationality. Suppose we do come up with a theory of the speaker's irrationality. Suppose we notice that the speaker's behavior is altogether rationally consonant with 'ka' meaning 'and', but that the speaker tends to make errors when forming complex statements involving some particular sentence. We might then hypothesize that it's difficult for the speaker to reason about that sentence—maybe it introduces a lot of cognitive load. This hypothesis once again lets us systematize the speaker's dispositions to assent: we expect that the speaker will assent to $\ulcorner r$ ka $s\urcorner$ if and only if he assents to r and to s unless either r or s is one of the sentences that causes cognitive load. The virtues will prefer a theory in which the agent is irrational if and only if the various irrational inferences the agent is disposed to make are patterned instead of piecemeal, and can be systematized into a theory of the agent's cognitive system that yields the patterns of irrationality. By positing various subpersonal cognitive mechanisms we explain the agent's dispositions, and this can go on without taking preservation of the agent's rationality to be any sort of goal.[12]

Sometimes it is argued that we should prefer models that assume humans to be rational because they are simpler than other models. Sober (1978) argues for this. Heil writes that it is useful to regard charity "as parsimony applied in the mental realm" (1994, p.120). It's not clear that these sorts of warnings put any additional strictures on scientific theory-construction. We already have parsimony in the mental realm: it goes by the name 'parsimony'. Moreover, we don't want to *equate* charity with parsimony, because we cannot guarantee from the outset that the most parsimonious (or otherwise virtuous) theory will be the one with the result that people

[12]The account has affinities with Cherniak (1986), who argues that we don't only holistically ascribe mental states and language meanings: we holistically attribute mental states and the meanings of our words along with a theory of the agent's cognitive system. This is in order to account for the ascription of irrational inferences that are the product of memory constraints and computational difficulty or intractability. Cherniak, however, takes his project to be one of individual psychological ascription rather than the ascription of our best scientific theory, and still thinks that a constraint of minimal rationality is needed on top of all this.

are rational. Thagard and Nisbett (1983) respond to Sober by presenting psychological evidence that people apparently behave irrationally in various domains; explaining away these apparent irrationalities will probably be less parsimonious than just building a model of irrationality in these domains. They present a moderate version of a principle of charity: "Do not judge people to be irrational *unless you have an empirically justified account of what they are doing when they violate normative standards*." I endorse this as a general methodological principle. However, I also note that it's easily derived from other norms of theory-construction. It's a corollary of epistemic conservativism. If you see possible counterevidence to the otherwise best theory on offer, don't immediately throw the whole theory out; wait to see whether there's an empirically justified theory that can better explain the recalcitrant data.

The reason that we usually ought not interpret people as irrational is that doing so would contravene other epistemic virtues. It is an *outcome* of interpretation that agents are rational—it is not a constraint on interpretation in the same way that parsimony is—just as it is an outcome of physics that there exist particles that have negative charge, not a mandate of anything like negative-charge constraint. Rationality is not needed as an additional interpretive constraint.

Why did so many great philosophers in the past posit rationality constraints if they are unnecessary, and why do so many supporters of minimal rationality projects today still posit similar constraints? Here are some possible contributing factors.

- Davidson thought that rationality went hand-in-hand with the anomalousness of the mental. In fact, Davidson used rationality constraints as a *premise* in one argument for anomalous monism: the "constitutive force of rationality" in the psychic realm has no equivalent in the physical realm.[13] However, one

[13] "My general strategy for trying to show that there are no strict psychophysical laws depends, first, on emphasizing the holistic character of the cognitive field [. . .] [I]n inferring [an agent's beliefs and motives] from the evidence, we necessarily impose conditions of coherence, rationality, and consistency. These conditions have no echo in physical theory, which is why we can look for no more than rough correlations between psychological and physical phenomena" (Davidson 1982a). Also see Heil (1989, p.574). Sellars and McDowell are other philosophers who lean heavily on the notion that mental ascription, justification, and rationality all belong to "the space of reasons." Dennett similarly distinguishes the intentional stance from the design stance and the physical stance partly by appeal to its commitment to norms of rationality.

can easily reject psychophysical laws without appealing to rationality: Putnam-style arguments that appeal to multiple realization are enough to establish the anomalousness of the special sciences. Being-a-predator is a multiply realizable property (there are many ways to be a predator), so there quite plausibly aren't any bridge laws that will link generalizations about predators to laws in physics. Yet we do not need to invoke anything like rationality constraints in order to explain boom-and-bust cycles in elk populations.

- Conflations between individual and scientific ascription might well have contributed to support for rationality constraints. See, for example, Dennett (1989) for an example of such conflation. The sort of explanatory gymnastics I went through when describing how a translation of ⌜p ka bu q⌝ might go is not a plausible account of what we do during actual mental state ascription. Bortolotti and Cherniak are probably right that we use heuristics if we do tacitly apply a theory when ascribing mental states, and a rationality assumption is a fairly plausible heuristic.

- Philosophers might have had the intuition that it is characteristic of a belief that it have a rational functional role,[14] and confused this with the thought that there must be a Principle of Charity or rationality constraint on interpretation. Lewis conflates these two conceptions in "Radical Interpretation" (1974). In this paper, he often describes the constraints as simply being our "general theory of persons" and he claims that it's an empirical question about whether people actually do have mental states that conform to our folk theory.[15] It is hard to square this with the *normative* principles he offers—principles that would *guarantee* rational beliefs—with the goal of reducing indeterminacy in

[14] I explore this intuition in the next section.

[15] "Karl might have no beliefs, desires, or meanings at all, but it is analytic that if he does have them then they more or less conform [to] the constraining principles by which the concepts of belief, desire, and meaning are defined" (Lewis 1974, p.335).

interpretation.[16] The principles can reduce indeterminacy only if they are *conditions* on radical interpretation, not postulates that can be empirically confirmed or disconfirmed.

- Rationality constraints might have been considered important in opposing eliminativism and defending realism about mental states. The claim that rationality has a constitutive force in mental state ascription implies a kind of unrevisability for psychology.[17]

3.3.4 The Error in Rationality Constraints

To summarize: we can concur with Bortolotti in thinking that there is no strong rationality constraint on belief ascription. However, Bortolotti took the non-doxasticist to be making an argument about individual ascription, and this was a misfire. Scientific ascription, not individual ascription, is the type of ascription relevant to the non-doxasticist's project. It is also the type of ascription that is discussed in the philosophical tradition that non-doxasticists tend to cite. Bortolotti's argument against rationality assumptions on *scientific* ascription comes down to her demonstrating various mundane but irrational mental states, and hoping that the non-doxasticist will agree that they are irrational beliefs. The non-doxasticist will do no such thing.

Rather, the reason that we should think that there are no rationality constraints on mental state ascription is that such constraints would be otiose. The rationality assumption was posited by philosophers to do work paring on down on a potentially unlimited number of mental state ascriptions, but other norms of theory construction already pare down interpretation enough. Moreover, even if an assumption of some sort were needed, the need for it to be a rationality assumption (instead of some other sort) seems unmotivated. We can therefore conclude that there is no

[16] "Karl should be represented as believing what he ought to believe…" (1974, p.336).

[17] The principle of charity originated from such concerns: Quine's appeal to the principle of charity was largely motivated by his desire to make classical logic unrevisable, which sat uncomfortably with his claims elsewhere that no scientific statement was immune from potential revision. See Haack (1996).

rationality constraint on the process of mental state ascription.[18]

3.4 Rationality and the Functional Role of Belief

3.4.1 Conceptual Analysis

There is a second way of understanding rationality to be a requirement of belief—one that doesn't make any claims about the process of ascription. One might think that the *functional role* of belief is one in which a belief stands in rational relations to behavior and other mental states. Let's draw another analogy with theories in physics. To be an electron, a subatomic particle must have certain features. It must have negative charge; it must have intrinsic angular momentum of ½, and so on. If some particle under observation does not display these properties, it isn't an electron. Similarly, for a mental state to play a belief-role, it might need to stand in rational relations with other mental states.

Unlike a rationality constraint on interpretation, this sort of requirement on the functional role of belief does not limit theories of psychology. Physics might eventually discover that there is nothing that plays the electron-role; we are not forced by

[18]Before leaving this topic, I should note that there is a third type of ascription—one in which rationality requirements are needed for reasons that are not purely epistemic or theoretical. Rationality constraints are sometimes thought to be important to the *metaphysics* of belief. Lewis, for instance, wrote that "It should be obvious by now that my problem of radical interpretation is not any real-life task of finding out about Karl's beliefs, desires, and meanings. I am not really asking how we could determine these facts. Rather: how do *the facts* determine these facts?" (Lewis 1974, p.333). This is a curious statement. 'Interpretation' or 'ascription' standardly refer to an epistemic process that generates beliefs about mental states, psychological theories, or other representational items. Lewis here denies that this is what he has in mind. Radical interpretation doesn't generate psychological theories. It generates psychological facts.

The claim seems to confuse epistemology and metaphysics. The world is indifferent to interpretations of it: Lake Erie either contains water or it doesn't, and the truth of this proposition is independent of any anything that would require a rationality constraint. Lewis, however, claims that a rationality constraint on interpretation is needed in order to reduce indeterminacy of the mental. What sense can we make of this claim if 'interpretation' doesn't refer to an epistemic process of theory construction? *We* might want to reduce indeterminacy in our theories to help settle our thoughts, but why does the *world* need an interpretive constraint to reduce indeterminacy?

Lewis, I believe, subscribes to a picture of metaphysics in which the virtues that make for a good theory are truth-conducive in their very nature—a theory is true *in virtue of* possessing theoretical virtues (Williams 2007). This is a variety of metaphysical pragmatism (Douven 2008). I think there are also reasons to read Dennett and Davidson as being tempted by this. It's arguable that for Dennett, being interpretable as a belief through an application of the intentional stance is *constitutive* of being a belief. I won't say much about this variety of mental ascription; the arguments I offer against rationality being a constraint on scientific ascription should work here, *mutatis mutandis*.

epistemic principles to hold onto a commitment to electrons at all costs. Similarly, psychology might discover that there is nothing that plays the rational-belief role.[19]

How do we know whether the belief-role really does require rationality? Conceptual analysis of BELIEF is often invoked. The methodology here is sometimes called the "Canberra Plan," associated with David Lewis and Frank Jackson. According to the Canberra Plan, philosophers are in the business of performing a priori analysis on folk concepts to determine what the world would have to be like in order for our common sense theories to be true. After the conceptual analysis is complete, only then do we consult empirical theory to determine whether there is anything in the world that fits our concepts (Braddon-Mitchell and Nola 2009).

There are three concerns about the application of this methodology in this case.

Firstly, there are familiar worries about the variability of concepts from population to population and even from within person to person (Stich 1988). There might not be such as thing *a* folk psychology with *a* concept of belief. Empirical work (that I do not think has been done) would need to be done. One would need to show that there is stability in the concept across persons and cultures.

Secondly, many of the proposals that claim that rationality is constitutive of belief do not seem to be borne of careful and sensitive conceptual analyses. The non-doxasticist who offers various necessary conditions of beliefhood typically proposes criteria that rule out many similar states that we would intuitively accept as beliefs.

For example, Graham and Stephens claim that the "propositional attitude interpretation of the notion of belief commits one, at least implicitly, to the following claims" (Stephens and Graham 2004): *(1)* beliefs possess representational content; *(2)* believers are confident or convinced that the representational content of the belief is true; *(3)* believers take account of the truth of the content or proposition in reasoning and action;[20] *(4)* beliefs tend to call up suitable affective responses or emotions, given one's values or desires.

---

[19]There are tricky semantic issues here about the referent of the ordinary term 'belief' if there is nothing that plays the rational-belief role; see Stich (1996) and chapter 3 of this dissertation.

[20]This is Graham and Stephens's formulation, but they should probably say that believers take account of what they *take to be the truth* rather than *the truth*.

(1) is unobjectionable, but the others are problematic. Regarding *(2)*: there are debates about how full belief is related to degrees of belief, but no one would want to say that full belief requires absolute conviction and confidence. If that were true, we would have very few beliefs, if any. Regarding *(3)*: the criterion simply makes it necessary that the content of a belief motivates action, but this is so weakly stated that delusions would satisfy the criterion (delusional subjects talk about their delusions, after all). Graham and Stephens clearly have something stronger in mind, but it is difficult to see how a stronger criterion would allow for irrational beliefs. Regarding *(4)*: not all beliefs seem to generate affect, and any affect that is the result of holding a belief seems incidental to the nature of belief. A psychopath who lacks empathy is not thereby barred from having beliefs about members of his family. It would be one thing if these were offered as features that were generally or usually true of beliefs, but they are meant to be criterial. Graham and Stephens claim that the doxasticist is committed to delusions having all of these features. If they allow that there are certain irrational beliefs that lack these features, their claim will not stand, for delusions could similarly be beliefs that lack these features.

Thirdly, and most importantly, the non-doxasticist is *revisionary*, and not particularly concerned with folk concepts. Remember: the folk are prone to calling delusions a type of belief (Rose et al. manuscript). Professionals are too, as can be seen from characterizations of delusions in diagnostic texts. Should this concern the non-doxasticist? I do not think so. The non-doxasticist hopes to provide an account of the deluded mind. We don't think that the folk concept of mass has much to tell us about physics; folk physics can get things radically wrong, and mature physics is allowed to innovate. So too should a mature psychology. The non-doxasticist might have to bite some bullets and say that superstitions (such as Nishad's superstition in Bortolotti's example) are not beliefs, but unintuitive claims are the price of scientific progress.[21]

[21] In chapter 5, however, I do argue that the notion of acceptance (the non-doxasticist state that I endorse) is implicit in folk psychology, and the folk do have a nascent understanding of it. I don't think science must be beholden to folk psychology, but the less revisionary an account is, the less it will worry a certain class of philosophers.

3.4.2 Toward Explanatory Arguments

What, then, is the non-doxasticist to do? Consider the philosopher who wants to argue that some other established cognitive state (such as imagination, memory, anger, or perception) is an irrational form of belief. Imagination isn't sensitive to evidence, but this is also true of irrational beliefs. Beliefs typically motivate action, but this is also true of imagination. Imaginings can be voluntarily formed, but this is not true of all imaginings (some are annoyingly unbidden), and it is arguable that some beliefs can be voluntarily formed.

The problem this philosopher will face is explaining why a certain class of "beliefs" behaves in such a systematically irrational way. I can, right now, voluntarily imagine that there is a dragon in the room. When I do so, I don't leap up and scream, and I don't revise beliefs that I have about the inexistence of dragons, but I do start forming inferences and beliefs about what it would be like if there counterfactually were a dragon in the room. By postulating a cognitive system in which imagination is a distinct attitude from belief, I can explain why the attitude reliably leads to other mental states. The philosopher who claims that imaginings are irrational beliefs will be unable to predict and explain (as I am able to) why this particular irrational belief (but not others) leads to counterfactual beliefs and yet largely does not otherwise affect non-verbal behavior.[22]

Attitudes must do explanatory work; there is no point in proposing one if it does not tell us anything that we could get by subsuming it into an already acknowledged one. Merely conceptual arguments for non-doxasticism cannot get off the ground. They must be accompanied by an explanatory argument giving a competitor theory. Bortolotti complains at one point that Currie's "(positive) thesis that delusions are imaginings [...] derives most of its plausibility from the (negative) thesis that delusions are not beliefs" (p.74). Perhaps she is right; perhaps much of the the force

[22] I am, of course, glossing over the theory explaining exactly how imagination affects behavior. For example, an adequate theory would need to explain why sometimes my dragon-imaginings don't lead to any overt behavior at all, and at other times they can lead to elaborate games of pretense, such as when a child and I jointly pretend that I, a dinosaur, am locked in combat with him, Batman. Currie and Ravenscroft develop the beginning of such a theory in their (2002).

of Currie's proposal comes from a (negative) conceptual argument. However, that is not the only source of its power. Currie also thinks that his positive thesis can explain what doxasticism cannot.

3.5 Summary

Conceptual arguments for non-doxasticism attempt to establish that delusions lack certain constitutive features of belief. The most popular conceptual argument—the irrationality argument—holds that delusions are not rational enough to qualify as beliefs. One type of irrationality argument holds that an assumption of rationality is required for mental state ascription at all, but this is mistaken. Scientific ascription does not need a rationality assumption, for our general tools of theory construction are sufficient. It is conceivable that a rationality assumption is required in individual ascription, on the other hand, but this type of ascription does not concern the non-doxasticist. A second type of irrationality argument uses conceptual analysis to reveal that rationality is necessary for belief. However, not only are these conceptual analyses usually perfunctory, they should not interest the non-doxasticist for the same reason that individual ascription should not. Non-doxasticism is revisionary, unconcerned with folk concepts and folk processes of ascription.

The onus is now on the non-doxasticist to flesh out a revisionary theory. In the next chapter, I discuss some contenders.

Chapter 4
Unsuccessful Non-Doxasticist Theories

The non-doxasticist needs to describe a benefit that is provided by refusing to count delusions that *p* as beliefs that *p*. This must involve providing not just a negative story about what delusions *aren't*, but a positive story about what delusions *are*. The non-doxasticist requires a competitor theory that can provide an explanation of the features of delusion that remain mysterious under doxasticism.

These competitor theories can vary greatly in how much they deviate from current philosophical theories and from commonsense talk about the mental. Bayne and Hattiangadi (manuscript) distinguish between two sorts of proposals for dealing with belief-like mental states: the conservative and the radical. Conservative accounts for explaining delusions subsume them within familiar psychological categories. Such accounts include:

- Accounts in which delusions are a currently-countenanced mental state that is not belief (such as an imagining or a piece of pretense).
- Accounts in which to have a delusion that p is to have a belief that q, where $p \neq q$.
- Doxasticist accounts on which delusions are non-paradigmatic beliefs.

Radical strategies for explaining delusions, on the other hand, attempt to develop previously unacknowledged mental state types. The radical non-doxasticist is in the business of *positing attitudes*.

In this chapter, I survey various proposals that have been offered, and argue that each is deficient in some way. The chapter has three sections, each dealing with a different sort of non-doxasticist strategy. The first concerns strategies that attempt to

explain delusions by arguing that their content is not what it appears to be. The second concerns strategies that posit mental states that are intermediate between belief and some other mental state, such as imagination. The third concerns strategies that claim that delusions are "framework propositions."

A theme that will emerge is that accounts which look radical—which appear to posit new attitudes—are in fact covertly conservative. Moreover, many offer no explanatory value over standard doxasticism; they are in fact merely terminological variants of doxasticism. We need to be more radical.

4.1 Redescription and Metarepresentation

4.1.1 Redescription Accounts

When a delusional individual asserts a bizarre proposition *p*, the doxasticist takes this to be the expression of a belief with content *p*. Non-doxasticists have more difficulty saying what is going on in this case. If it's not a belief in *p* that generates the assertion, then what does? One option is to posit that the person bears a mental attitude other than belief toward *p*. However, some non-doxasticists find no need to break with the conventional stock of psychological attitudes, and instead toy with the *content* rather than the attitude. Let's call these non-doxasticist strategies *redescription strategies*. A redescription strategy attempts to sever the explanatory link between a subject's utterance with apparent content *p* and a mental state that *p*. This can be done by either claiming that *(1)* the delusional utterance does not actually express its apparent content *p*, or that *(2)* the person's assertion has content *p*, but we can explain the assertion by appealing to the person's belief that *q*, where *q* is distinct from *p*.

There are three main forms of the redescription strategy in the literature. The first is to adopt the "senseless ravings" and "empty speech act" views of Berrios and Jaspers that were discussed in the first chapter. To recount: Jaspers held that ascribing mental states to a person required being able to "genetically understand" their

words; the psychiatrist must be able to make sense of the mental states in the person's biographical life story. Because the madman is entirely beyond understandability, his words are mere "senseless ravings" (Jaspers 1963, p.577). The psychiatrist German Berrios followed Jaspers in holding that delusional speech is senseless, arguing that the reports of the delusional individual are "empty speech acts that disguise themselves as beliefs" (Berrios 1991, p.8). On these views, a subject does not report a belief that *p*, because his or her utterances fail to express any content at all.

"Senseless ravings" is overly harsh, but it is not an altogether inaccurate description of the utterances of some individuals with mental pathologies. Schizophrenia is often accompanied with thought disorder and dysfunctional speech; the individual produces semantically garbled language. The "word salad" produced by individuals with thought disorder often does not look like it expresses a belief at all. For instance, Covington et al. (2005) write,

> Very florid impairments are sometimes reported. When a patient with schizophrenia says something like:
>
> > Oh, it [life in a hospital] was superb, you know, the trains broke, and the pond fell in the front doorway. (Oh et al., 2002, p. 235)
>
> do the words mean anything at all? Is the patient actually expressing a thought of a pond falling in the front doorway?
>
> Perhaps not. (Covington et al. 2005, p.14)

Similarly, Louis Sass has recounted to me a person who, when asked to explain the proverb "Don't swap horses when crossing a stream," replied, "That's wish-bell. Double vision. It's like walking across a person's eye and reflecting personality. It works on you, like dying and going to the spiritual world, but landing in the Vella world." Do we really want to say that person entertains a thought with that content? Again: perhaps not.

However, it is untenable to claim, as Berrios does, that *all* expressions of delusion have *no* semantic content. Many delusional individuals are perfectly capable of holding conversations about their delusions. They draw inferences. If delusion reports are not belief reports, some other explanation must be offered to explain these

semantic capabilities. The utterances above are not the best models of delusion; it is standard to distinguish delusions from formal thought disorder, and the dialogues above are expressions of thought disorder. Though the two are often comorbid, delusions can exist without thought disorder, as they do in many monothematic delusions brought about by organic brain injury. This suggests that we have two separate phenomena at hand, even though in schizophrenia they are often found together and presumably have a common etiology. Jaspers and Berrios may be misled in treating the word salad that arises from thought disorder as exemplary of delusional speech.

A second strategy is to interpret the delusional subject's assertion *metaphorically*. "Surely he can't *literally* mean that!" is a common reaction to the crazed assertions of the schizophrenic. This has led to the thought that patient must be trying to express something else, confusedly or badly. Why do delusions not act like normal beliefs? According to the alternate-content theorist, this is a trick question: they do act like normal beliefs, but the content of the belief is not immediately apparent.

Metaphorical reinterpretation of delusional assertions has been historically popular. For instance, psychodynamic and psychoanalytic theories have long maintained that our psychologies are largely controlled by subconscious motivations and repressed memories of the primal stage. This hidden inner life burbles out into dreams, slips of tongue, and other scarcely-noticed behaviors, but it only does so encoded in symbol and archetype. A skilled interpreter is needed to decrypt the metaphorical messages. This Freudian picture of the mind has surely been more influential than it merits, but the suggestion that delusional patients communicate metaphorically is widespread. R. D. Laing (1969) argues that delusional patients are metaphorically expressing their emotions. Richard Bentall (2003, p.29) recounts a patient of his research assistant's who claimed that she had been turned into a portfolio. This at first seemed to him to be a perfect instantiation of Karl Jaspers's famous claim that psychotic patients are ununderstandable. How could she literally think she was a portfolio? Eventually, however, his research assistant came to learn that this patient had been diagnosed with a rare disease and had been subjected to

batteries of physiological examinations. Bentall's assistant concluded that her assertions metaphorically reflected the dislike she had of being uncaringly interpreted as an object of scientific study. This sounds like a plausible interpretation.

Similarly, the Cotard delusion is typically comorbid with severe depression and can be alleviated with antidepressants. It seems possible that the patient who says "I am dead" might be hyperbolically communicating that they are unimportant, or that they are unable to feel emotion, or that they wish to be dead, or something of the like. Louis Sass (2004) points out that we all use the word 'dead' to signify that someone is exhausted or emotionally distant. Why not think that the delusive patient is also being creative with his words? Moreover, patients with forms of brain damage that render language use difficult will often resort to metaphor. Subjects with Alzheimer's disease or some forms of aphasia, unable to call up the word they are looking for, can hit upon forms of communication that seem out-and-out poetic. It would not be surprising if for some reason cognitive confusion or performative disability rendered psychotic patients unable to express what they really wanted to express, and they were forced to communicate in a roundabout way.

It is certainly true that many of the outlandish claims of the delusive might be chalked up to metaphor. We need to be careful and sensitive in interpreting their words. However, the metaphorical account will not suffice to fully explain all the oddities of delusions. There is a world of difference between the poet and the madman. Compare the delusional patient who claims he is made of glass to the romantic who, overtaken by a sense of emotional fragility, claims the same. If you ask the poet why you can't literally see through him, he'll admit that you can't literally do so, but also claim that this isn't in conflict with his earlier poetic claim of being made of glass. He was, after all, just being metaphorical. Delusional patients will feel the conflict, and will either confabulate some justification that makes the two claims compatible, or shut down the conversation. More important is the fact that delusional patients, in certain contexts and in certain ways, act as if their delusions *are* literally true. Although people treat their delusions metaphorically in some ways, they don't *entirely* treat their delusions as metaphorical. For example, consider delusional parasitosis:

the delusion that bugs are crawling under one's skin, sometimes experienced in response to cocaine withdrawal. When I am itchy, I might describe the sensation by saying that I have bugs under my skin. I certainly will not call up exterminators, as sufferers of delusional parasitosis might do. Some story must be told about how delusional subjects can treat their delusions as literally true in some respects but not others. Moreover, there are many delusions that are very hard to interpret even metaphorically.

The third redescription strategy—the most well-developed—is employed in the *metarepresentational* accounts most forcefully argued by Currie and his various coauthors (Currie and Ravenscroft 2002, Currie 2000, Currie and Jones 2006, Currie and Jureidini 2001).[1] Instead of interpreting the delusional subject's reports metaphorically, they ought be interpreted metarepresentationally. The schizophrenic should not be said to believe his delusion that *p*; he imagines that *p*, and because he misidentifies this imagining as a belief, he comes to believe that he believes that *p*.

4.1.2 Currie's Metarepresentational Account

The metarepresentational gambit that Currie employs is not unique to the literature on delusions. It is a tactic commonly used by those who subscribe to rationality constraints in order to explain apparently irrational beliefs. For example, Lewis is committed to radical interpretation guaranteeing that subjects accede to all logical truths, and Davidson is committed to any individual having mostly true beliefs about the world. What does Lewis say about the dialetheist, who claims to believe in true contradictions, and what does Davidson say about the skeptic who claims not to know that the external world exists? Lewis and Davidson should think that, if the dialetheist and the skeptic are sincere in their philosophical commitments, then they are "deluded" in the vernacular sense of the term. The dialetheist *thinks* he believes contradictions and the skeptic *thinks* he does not know that the world exists, but

[1] For brevity's sake, I'll only mention Currie when discussing proponents of the metarepresentational theory, as he is the common link who has produced multiple publications defending the view. You may mentally inscribe an '(and coauthors)' after each mention of his name.

they are both sorely mistaken about what they actually believe. Their behaviors belie their philosophical avowals. The skeptic goes on interacting with the world, publishing papers, giving talks, and so on. These actions are indicative of his real beliefs about the world. Similarly, Stalnaker's diagonalization strategy for dealing with Frege cases and problems of logical omniscience can also be interpreted as an instance of the metarepresentational gambit (Stalnaker 1984). Someone who asserts "all primes are odd" can't actually believe that all primes are odd, for it's a necessary falsehood. Rather, they believe that the sentence 'all primes are odd' expresses a truth. The content of this belief is false, but only contingently false. In worlds in which 'primes' means SUMS OF TWO CONSECUTIVE NATURAL NUMBERS, the belief is true.

Metarepresentational strategies have lately been popular as a way of explaining and naturalizing religious belief.[2] It is problematic for various reasons to think that self-proclaimed religious folk genuinely believe their avowals in the existence of God. The best explanation is that they don't believe; they just believe they believe. Dennett (2007) and the economist Steven Landsburg (2009) argue that no one would be so irrational that they would cheat and steal and lie if they really believed in eternal hell. They take a metarepresentational approach in order to resolve the tension. Rey argues that self-identified believers don't have the emotional response to death that one would expect if they genuinely did believe in God (Rey 2007). Sperber (Sperber 1996, 2000) posits metarepresentational beliefs in order to explain how the cultural transmission of mysterious and only partially-understood propositions is possible. A person who asserts a sentence such as 'The Father, the Son, and the Holy Ghost are One' might actually believe a metarepresentational proposition, such as " 'The Father, the Son, and the Holy Ghost are One' expresses a holy mystery," or "My priest tells me that 'The Father, the Son, and the Holy Spirit' is true."

Currie's account is in this ballpark. According to Currie, the delusional person

[2]See, for instance, Sperber (1996, 2000), Atran (2002), Boyer (1994), Dennett (2007), Landsburg (2009), Rey (2007).

imagines a proposition *p*, but a deficiency in self-monitoring causes him to misidentify this imagining as a belief. So, the person doesn't genuinely believe *p*—he imagines *p*—but he believes that he believes *p*. In presenting the advantages of his theory (Currie and Ravenscroft 2002, p.176–9), Currie points out that schizophrenic patients do not always act on, or pretend to act on, the contents of their delusions. And he notes that they do not often draw correct inferences from their delusions or attempt to reconcile them with other beliefs. These are features that we would not be surprised to see in imagination. In fact, the disordered train of thought we encounter in schizophrenia would not be thought shocking if seen as the flow of imagination. He gives an example of a soldier, who upon hearing cows lowing, becomes convinced that he is about to become slaughtered like cattle. It would not be incredibly unusual for a person to idly daydream this when prompted by the lowing, suggesting that there may be a link between the formation of the delusion and the formation of an imagining. However, delusions are not simply imaginings. Where the behavior of delusional individual deviates from that of the dreamer, we can appeal to the delusional individual's metarepresentational belief. Why does the Capgras patient say that his wife is an imposter if he is only imagining that she is an imposter? Because he thinks he believes that she's an imposter.

Particularly compelling is Currie's description of thought monitoring, and his explanation of how the imagining is mistaken for a belief. It appeals to our current best models of disorders of volition, such as passivity phenomena and alien hand syndrome. Schizophrenics will often claim that their bodies are being controlled by someone else, or that their thoughts are not their own. The reigning explanations of these phenomena is owed to Christopher Frith; it appeals to the 'efference copy' model of volitional action (Frith 1992, Frith et al. 2000). Once a person forms an intention to act in a certain way, a command is sent to motor control. An *efferent copy* of that same command is sent to a *comparator* and stored there. When we receive information about the way that our body moves (through proprioception, vision, and other senses), a representation of that bodily movement is compared to commands stored at the comparator. If there is no efferent copy with matching content, then

the movement is experienced as not controlled by the agent and non-volitional. This is why we experience a felt difference when we raise our arm and when our arm is raised for us.

In schizophrenics suffering from an alien hand, there is something wrong with the comparator. When the schizophrenic sends a command for his hand to move in a certain way, and his body actually does move that way, the comparator should say: "yes, that all checks out!", and generate a volitional experience. But because the comparator is broken, it does not register a match, and so the person feels pushed around. This sensation of willing is allegedly generated not just when we move our bodies, but also when we think. (Currie approvingly cites the notion that thinking is a form of motor action.) Schizophrenics don't feel that their thoughts are their own because of a broken comparator.

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.

The theory is ingenious. Let's criticize it.

4.1.3 Criticism of Currie

Imagination has long been summoned in explanations of delusion formation. Here is Locke, in his *Essay Concerning Human Understanding*:

> Madmen "do not appear to me to have lost the faculty of reasoning, but having joined together some ideas very wrongly, they mistake them for truths; and they err as men do that argue right from the wrong principles. For, by the violence of their imaginations, having taken their fancies for realities, they make the right deductions from them. Thus you shall find a distracted man fancying himself a king, with a right inference require suitable attendance, respect, and obedience;

> others who have thought themselves made of glass, have used the caution necessary to preserve such brittle bodies. Hence it comes to pass that a man who is very sober, and of a right understanding in all other things, may in one particular be as frantic as any in Bedlam. . . . In short, herein seems to lie the difference between idiots and madmen; that madmen put wrong ideas together, and so make wrong propositions, but argue and reason right from them; but idiots make very few or no propositions, and reason scarce at all."[3]

The theory Locke describes here is very similar to Currie's, but it does not have a metacognitve component. Instead of the delusional subject misidentifying their imagining as a belief, they *use* their imagining to form new beliefs, as if it were a belief that could legitimately act as an acceptable premise. Reasoning is otherwise unimpaired. Let's call this *the Lockean theory* of delusion formation.[4] According to Currie, an individual might imagine that, say, there is a power plant on the inside of his body, and go on to believe that he believes that there is a power plant inside his body. On the Lockean view, a delusional individual who imagines that there is a power plant inside his body will then infer whatever he would infer if this were a belief (for example, that his insides are radioactive).[5] What is wrong with this theory? Why not think that the content p of the imagining is used as a premise rather than positing that the metarepresentational content $\ulcorner$I am imagining $p\urcorner$ is used as a premise?

Currie's metarepresentational theory has certain explanatory benefits: the story about the comparator model tells us the origins of the delusion. However, the Lockean can exapt most of Frith's story about the comparator for his own purposes. It might well be that when an imagining is experienced as non-volitional, we don't form a meta-belief about the kind of mental state that it is; we instead *use it as* another sort of mental state in reasoning. Perhaps when an errant thought runs through

---

[3]It is somewhat interesting that Descartes (as mentioned in chapter 1) and Locke both refer to the delusions that one is a king or that one is made of glass. I haven't been able to determine their source, if there is one. One suspects some slight cribbing going on.

[4]McGinn (2005) is a modern proponent of a similar view.

[5]The theory leaves open whether the content of the imagining would also be believed. One could hold that, because inferring p from p is rather useless, it is not something we naturally do all that often, and the delusional subject wouldn't necessarily infer the belief that p from the imagining that p. In this case, the account might be considered semi-doxasticist—to be deluded that p would be to imagine that p and use the imagining as a premise in reasoning.

our heads and isn't identified by the comparator police as an imagining, it just gets straightaway slotted into our knowledge store.[6]

In any event, the Lockean theory has a problem. It does not explain the troubling data that we want a theory of delusions to explain. Locke explains how delusional beliefs are formed, but the theory doesn't handle the stickiest features. The theory doesn't explain why delusional individuals retain the bizarre beliefs instead of subsequently rejecting them; it doesn't explain double-bookkeeping; and most importantly, it doesn't explain the circumscribed *effects* of delusions on action and theoretical reasoning.

Does Currie fare any better? He seems to think so: he claims the metarepresentational addition of his theory can handle these features, and this is an advantage of his account. His reason is that imaginings have the features of delusion that are most troublesome: they are not sensitive to evidence, they do not typically alter action, or other belief, and so on: "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).

The problem here is that as the theory was presented above, delusions are not imaginings: they are metarepresentational beliefs *caused* by wayward imaginings. Currie is not consistent on this point: he waffles on what he wants to affix the term 'delusion' to. At some times he says that delusions are imaginings; at other times he says that delusions are misidentified imaginings; at other times he says that they are metarepresentational beliefs. In itself, this is not a problem: this might just be a case of Currie having a non-substantive terminological dispute with himself (as it were) about the proper referent of the term 'delusion'. It's perfectly possible to describe Currie's theory without using the term 'delusion.' A schizophrenic person imagines *p*; owing to a faulty comparator he identifies that imagining as a belief, and thereby

[6]Recall the Wittgenstein thought that the difference between perception and imagination consists in the feeling of will. There is nothing in this theory about metarepresentation; that was an addition of Currie's.

comes to believe that he believes *p*. It should now be clear that Currie cannot appeal to the circumscription of imagination to explain the circumscription of delusion. He needs to appeal to the circumscription of metarepresentational beliefs.

We now face a tricky question about whether metarepresentational beliefs are circumscribed, and whether they are circumscribed in the same way that delusions are circumscribed. How should we expect someone to behave when they believe that they believe *p*, even though they don't actually believe *p*? Currie's response is that they will say *p*, but not otherwise act as if *p*. Our actions manifest what we believe, and our words manifest what we think we believe. I maintain that this is not a good explication of a belief in a belief that *p*. Beliefs have non-verbal manifestations, and metarepresentational beliefs are beliefs too; we should expect them to also have non-verbal manifestations. These are manifestations that Currie ignores.

Consider that not all metarepresentational beliefs are expressed primarily in verbal behavior. My beliefs about *other people's* beliefs are not entirely manifested in verbal behavior. They are manifested in the ways that respond to others and coordinate my actions with others. Take, for example, the following scenario: I think that my partner has forgotten her dentist appointment, so I leave a reminder on her laptop. Doing so manifests a bunch of my metarepresentational beliefs about what she believes. I believe that she believes her laptop is on her desk, that she believes that she needs her computer, that she does not believe she has a dentist's appointment, etc.

Now, suppose that I am worried that I will forget about *my* dentist appointment, so I leave a note for myself on my laptop. Doing so similarly manifests beliefs that I attribute to my future self: my future self believes that his computer is on his desk, etc. It's fair to suppose that I attribute beliefs to my future self based upon what I currently take my beliefs to be. The reason I think my future self will believe that his computer is on the desk is that I think that I currently believe that my computer is on my desk, and I don't expect these beliefs to change in the interim. Thus, just as it is possible to manifest attributions of belief to other people non-verbally, it is possible to manifest self-attributions of belief non-verbally.

Metarepresentational beliefs are often manifested in the construction and initiation of *plans*. In order to create a complicated plan, we need to consider how we would act in various counterfactual scenarios, and this will be influenced about what we believe we actually believe. Let's say that my drill breaks and I need to buy a new drill bit. I decide that when I go shopping for groceries later in the day, I will travel down Avenue B instead of my more common Avenue A, because there's a hardware store on Avenue B and I'll be able to run in and get a drill bit. My implementing this plan does not only manifest a belief that the hardware store sells drill bits; it also manifests a belief that I believe that the store sells drill bits.[7] If I did not attribute this belief to myself, I would not expect to successfully buy any drill bits at the store. After all, if I were to give this plan to another person—if I asked someone to swing by the hardware store because I destroyed my drill bit—I would be manifesting a belief that the other person believes that there are drill bits available for purchase. Not much changes in the case where I give the plan to myself.

The question now is whether a delusional individual's plans—and not just verbal affirmations—reveal a metarepresentational belief. I do not think that they do. Suppose I believe that my friend believes that his wife has been replaced with an evil robot. (I'm not attributing a delusion to him in this scenario, mind you; I'm attributing an actual belief.) I could not expect him to successfully carry out a plan that involves him bringing a drill bit back to his house. After all, if he really believes that his wife has been replaced with an evil robot, he wouldn't go anywhere near that house! Now, suppose a person with the Capgras delusion breaks her drill bit, and begins to formulate a plan to fix the drill. If she really believed that she believed that her

[7] Please be aware that I do not claim that this metarepresentational belief must occurently enter into my thoughts. It might well be the case that when I develop plans of action, in order to determine how I would act in various situations, I use simulation—I imaginatively inject myself into various counterfactual scenarios. This would mean that I would not have to explicitly reason with the premise, 'I believe that there are drill bits available at the hardware store'. However, one needn't have a proposition enter into conscious thought in order to be the object of belief. Rather, the construction and implementation of all but the most simple plans will *manifest* a metarepresentational belief. My acting on the plan to acquire a drill bit also manifests the dispositional belief 'I will not be eviscerated by dinosaurs if I go to the hardware store', even though that proposition does not enter into my explicit reasoning. (Usually.)

husband had been replaced with an evil robot—and that was the only thing pathologically unusual about her—then she would never expect to come home. She would expect to run away and call the police, or whatever else you'd expect from a person who genuinely thinks that her husband has been replaced. She couldn't put that plan in motion. However, individuals with the Capgras delusion are able to live with the purported imposters, and carry out relatively normal lives with normal plans.

Currie notes that verbal and non-verbal behavior often come apart. His explanation of this phenomenon, which is a typical one, is to claim that we verbally express metarepresentational beliefs and act on non-metarepresentational beliefs. However, this does not sufficiently do justice to the role that metarepresentational beliefs would have in our reasoning and our lives. Currie is here falling prey to a temptation that many have felt when attempting to characterize belief. It's tempting to overemphasize the role of verbal affirmation in belief, and to think that *all* beliefs (and not just metarepresentational beliefs) are primarily manifested verbally. It's tempting to take belief to simply be a disposition to sincerely affirm a proposition. However, this is usually seen as an overly linguistic characterization of belief.[8] Beliefs are manifested in our intentional actions, in our reasoning, and in the ebb and flow of of our other mental states. Verbal behavior is just one out of many manifestations of belief, and metarepresentational beliefs are simply beliefs with a certain sort of content.

In addition to the criticism that Currie does not accurately portray the functional role of metarepresentational beliefs, one can also criticize his portrayal of the functional role of delusions. Currie identified metapresentational beliefs and delusions by appealing to their both being manifested verbally. However, it is not quite right to say that delusions are only manifested verbally. Delusions are only *partially* behaviorally and affectively circumscribed, and they do have *some* effects on non-verbal reasoning and behavior. This might seem like it would help Currie out—I just argued that metarpresentational beliefs have a broader role than affecting verbal behavior—but it is hard to see how a metarepresentational belief will explain the precise *pattern*

[8] In the next chapter, I call this the 'assent theory of belief', and subject it to further criticism.

of circumscription that we see in genuine delusions. A person who says he is Jesus might dress as Jesus does while also being unperturbed about his residence in a mental hospital. Is there any way to interpret those inconsistent patterns of behavior as a metarepresentational belief that he is Jesus? Currie has an explanation. He offers that metarepresentational beliefs will eventually get entrenched as non-metarepesentational beliefs; he writes,"it may well be that over time certain delusions come to take on the action-guiding character of beliefs." But as Egan points out (2009, p.11), this means that we *should* expect delusional persons to act just like imaginers (and believers in the metarepresentational proposition), and then at some point in time, see them switch to acting as true believers. We should never expect to see the "somewhere-between-belief-and-imagination pattern of circumscription that we actually see." Egan, and Bayne and Pacherie too (2005), also argue that Currie's theory makes some other false predictions. If a person has a merely metarepresentational belief, then we would expect a person to act more delusional when in a reflective mood, as she attends to beliefs about her mental state. This doesn't appear to be the case.

Although Currie officially adheres to a metarepresentational account of delusion, he does float a somewhat different idea in a recent coauthored work, perhaps recognizing that he equivocates between treating delusions as imaginings and as metarepresentational beliefs:

> we suggest—rather tentatively—a somewhat different view: that delusions considered as a class of states do not fit easily into rigid categories of either belief or imagination. 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 Jureidini 2001)

They do not elaborate, but this is a more radical type of non-doxasticism than that in Currie's previous work. It bears many similarities with the "bimagination" view that Egan has recently proposed. This will be the topic of the next section.

4.2 Bimagination and In-Between Belief

4.2.1 Egan's Bimagination Account

In "Imagination, Delusion, and Self-Deception" (2009), Egan proposes that delusions are not straightforward beliefs. Rather, a delusion is an instance of a novel attitude that is somewhat intermediate between imagination and belief: "bimagination." Bimagination has some of the distinctive features of believing and some of the features of imagining. For example, delusions, like imaginings, are not very sensitive to evidence, and like imaginings, they are inferentially and behaviorally circumscribed. However, their behavioral effects are more profligate than those of standard imaginings—they share many of the dispositional features of belief as well. Bimagination is presented as a response to Currie: it accounts for the seemingly intimate relation between delusion and imagination without being a metarepresentational theory. As it turns out, Egan's theory offers us an apparently radical form of non-doxasticism that turns out to be less radical than it first seems.

Egan's primary goal in the paper is to make conceptual space for the notion of bimagination rather than to give a positive argument for delusions being instances of it. Thus, he spends the bulk of his paper arguing that there is room in our cognitive theories for hybrid attitudes: attitudes that are neither fish-nor-fowl—neither belief-nor-imagining—but something in between. Egan imagines that some opponents of bimagination might vouch for a restrictive view on mental attitude types, according to which there are only a few roles available, and every attitude of a particular type has the same functional profile. The existence of *belief fragmentation* challenges this assumption.[9]

Egan presents a picture of *fragmented belief*, or *compartmentalized belief*, or *partitioned belief*, in which the behavior-guiding role of a belief is active only in certain

[9] I produce an argument with a similar conclusion in the next chapter that does not rely on belief fragmentation.

contexts, or the belief is only able to affect certain domains of behavior.[10] A paradigmatic belief that *p* disposes the believer to act all the time and in every respect in ways that would likely be successful if *p*. Yet, most of our beliefs fail to be paradigmatic. In many contexts I will be disposed to act as if I believe *p*, but there will almost surely be contexts in which my dispositions are not those of a paradigmatic *p*-believer. For example, Egan offers the folk psychological distinction between *recognition* and *recall*. Recalling a fact is more difficult than recognizing that the fact is true: a source of frustration for *Trivial Pursuit* players everywhere. If you ask me the capital of Peru, I might draw a blank. If you then tell me it's Lima, I'll smack my head and say "I knew that." In contexts in which I'm asked to *recall* the capital of Peru, my behavior is that of an ignorant person, but in contexts in which I'm asked to *verify* whether Lima is the capital of Peru, my behavior is that of a knowledgeable person. When it comes to the functional profile of the belief "Lima is the capital of Peru," some of my reactions swing one way, and some of my reactions the other.

There's nothing especially contentious about this. It is common for people to have inconsistent beliefs, or to fail to bring pieces of knowledge to bear upon one another, or to fail to recall pieces of known information in certain contexts. The best and most natural explanation of these phenomena is that only some of the representations in our heads are accessed, or are available for access, by any particular subsystem at any particular time. Only certain representations are *currently active* at any one time. Cherniak (1986) describes a man who met his end by lighting a match to peer into a gas tank. The poor man presumably believed that it is dangerous to light flames around gas fumes, but he unhappily failed to activate that belief and have it influence his plans.

Beliefs are often characterized in something like the following way: agents are disposed to acts in ways that would satisfy their desires if their beliefs were true. A picture of fragmented belief might include something about this only being true for currently active beliefs. As a first pass: agents are disposed to act, in a context *c*, in

[10] Egan gives a fuller account of his thoughts on fragmentation in his (2008). Also see Davidson (2001), Stalnaker (1984), and Lewis (1982).

ways that would satisfy their c-active desires if their c-active beliefs were true.[11]

The problem with relativizing fragmented beliefs only to contexts is that it implies that in certain contexts we will act like we would if we paradigmatically believed p, and in other contexts we will act like we do not. Descriptions of fragmentalized or compartmentalized belief will often use temporal language such as 'sometimes' or 'currently active': e.g. an agent with the fragmented belief that p sometimes behaves as if he believes p, and at other times he does not. The problem is that we often behave as if we believe two inconsistent things at a single time. Self-deceivers, for instance, will say one thing but do another. Or, consider the Ebbinghaus optical illusion: a subject will verbally report that various circles appear to be of different sizes, but when asked to manipulate the circles by using her hands, her movements are not those of someone who is taken in by the illusion. Verbal and non-verbal motoric behavior come apart: verbal behavior indicates the presence of a certain belief but motoric behavior does not. It appears that it is not just the case that particular representations are active in particular *contexts*. Each representation is only able to affect a particular *domain* of behavior.[12] Of course, it is also difficult to individuate or characterize contexts.

Thus, Egan's final characterization of the behavior-guiding role of belief and desire—one which allows for fragmentation—is that "agents are disposed to act, in a context c and within a domain d, in ways that would satisfy their $\langle c, d\rangle$ active desires if their $\langle c, d\rangle$ active beliefs were true" (2008). Fragmentation demonstrates that the functional role of a belief has various "bits" or "elements" to it, and only some of these bits may be present in certain contexts. Once Egan has established that lapses in inferential reasoning that can be chalked up to fragmented belief, he hopes to have eroded his reader's hesitance at accepting the idea of states that are intermediate

[11] Egan also thinks that desires can be fragmented like belief.

[12] It's difficult to say exactly how to characterize or individuate domains of behavior. It's too coarse-grained to simply count our verbal capacities and non-verbal motoric capacities as responsible for two distinct domains, for instance. Cases in which verbal and motoric behavior are especially easy to come up with because they are so overt, but it is not hard to think of entirely non-verbal behavior that indicates two fragmented beliefs are simultaneously active, such as when one puts on one's glasses to be able to better look for one's glasses.

between beliefs and imaginings.[13]

4.2.2 Schwitzgebel's In-Between Account

Schwitzgebel offers an account of delusions in much the same vein. He argues for the existence of mental states that are intermediate between the more well-known attitudes, and calls these 'in-between beliefs'. In a series of publications, Schwitzgebel (2001, 2002, forthcoming) relies upon them to explain an impressive array of mental phenomena, including partial forgetting, self-deception, confidence judgments, and delusions. This characterization of in-between belief differs from Egan's bimagination in the specifics, but the basic idea is the same: belief boxes have fuzzy boundaries, and there are mental states that have some, but not all, of features of prototypical beliefs. Once we admit this possibility, we can call upon these states to explain otherwise inexplicable behaviors, including those of psychotic individuals.

Schwitzgebel's is a dispositional theory of belief.[14] According to his theory, belief-that-*p* is a 'dispositional stereotype': a stereotype made up of various dispositional properties. For Schwitzgebel, an object has a dispositional property if and only if a certain type of subjunctive conditional is true of it. The conditional must be of the form: in condition *C*, object *O* will enter or remain in state *S*. Beliefs can't be *single* dispositional properties (according to this conception of dispositional properties), because beliefs cause persons to act in all sorts of different ways in all sorts of different circumstances. Your belief that there is beer in the fridge is marked by a disposition to look in the fridge if one wants beer, to tell one's friend that there is

[13] "The goal here is just to open up a space for saying that there are such representations and such representational states by undermining the all-or-nothing-roles view that, if we endorsed it, would rule out such intermediate roles, states, and representations" (2009, p.270); "[I]t's not the case that anything that's got one element of a certain package has also got to have all of the rest, since we see a variety of mix and match patterns even within belief" (2009, p.274).

[14] His view is pretty similar to standard dispositionalism, with one modification. Dispositionalism holds that mental states are (or realize) dispositions to go from certain inputs (stimuli and other mental states) to certain outputs (behaviors and other mental states). Schwitzgebel holds that there might be phenomenal mental states that cannot be described in this way. Thus, on his view, some mental states are phenomenal and some are dispositional. Dispositional mental states are still characterized the same way: they are dispositions to go from stimuli and other mental states to behavior and other mental states, but 'mental states' in this characterization refers to non-dispositional, phenomenal states as well.

beer in the fridge if he asks for one and you are feeling generous, and so on. Beliefs are *collections* of dispositional properties.[15]

To believe that p is simply to match a belief-that-p stereotype to an appropriate degree and in appropriate respects, and these will vary contextually. An agent has an *in-between* belief-that-p if some, but not all, of the dispositions in the belief-that-p stereotype hold of the agent. Belief-that-p is, thus, a vague or indeterminate matter, and delusions lie in the penumbral zone. Tumulty (2011, 2012) follows Schwitzgebel in arguing that delusions are in-between beliefs that play some, but not all, of the dispositional roles associated with the folk-psychological notion of belief.[16]

How does Schwitzgebel's account compare to Egan's? Bimagination and fragmented belief can both be characterized as sorts of in-between belief. A fragmented belief-that-p is a belief-that-p that will generate behavior in domain d when it is active in contexts c, but that is not active in all contexts and/or does not generate behavior in all domains. Therefore, it can be characterized with a subset of Schwitzgebelian dispositions within the stereotype of a belief-that-p: all of those that have c in the antecedent and d in the consequent of the corresponding subjunctive conditionals. It is, in other words, a state that behaves like a belief-that-p, but only sometimes and in some respects. Bimagination is a state that has many of the dispositions in the stereotype of belief, and some of the dispositions in the stereotype of imagination.[17]

---

[15]Schwitzgebel assumes that all dispositional properties are single-track dispositions (i.e. dispositions to manifest in a particular way under a particular stimulus), and that multi-track dispositions (i.e. dispositions that would manifest differently in a variety of stimulus conditions) are actually collections of single-track dispositions. This claim is contentious—there are those who think that multi-track dispositions are "Lockean powers" that cannot be reduced to collections of single-track dispositions. I myself think that Schwitzgebel's reduction is subject to too many counterexamples to be successful, but I won't follow this line of thought here. I mention it only to flag Schwitzgebel's theoretical commitments.

[16]Tumulty, however, holds that an adherence to the norms of rationality is the characteristic feature of belief's dispositional role: as mental states fail to adhere to the epistemic rules that govern belief, they become less determinately beliefs. Schwitzgebel does not mention rationality requirements, instead focussing on the behavioral and dispositional effects of belief. As I have dealt with rationality requirements in the previous chapter and do not find them necessary for belief ascription, I'll continue to focus on Schwitzgebel rather than Tumulty.

[17]I do a little damage to Egan's view here: the way Egan puts his view does not commit him to Schwitzgebel's reductionism about dispositions. Schwitzgebel reduces multi-track dispositions to collections of single-track dispositions, and attributions of single-track dispositions to attributions of subjunctive conditionals. Egan could be a Lockean realist about dispositions and think these sorts of reduction are not possible.

The apparent difference between Egan and Schwitzgebel is that Egan proposes a new propositional attitude. Bimagination is presented as a newly-discovered mental state. In-between belief however, is not meant to designate a natural or well-formed category. As Tumulty writes, Schwitzgebel is "not proposing in between belief' as a new ascriptive choice" (2011, p.616). The term is is simply meant as a way of referring to those wayward dispositional complexes that share some characteristics with belief, just as the term 'noodle-like object' refers to a hugely heterogenous variety of objects that share certain characteristics with noodles. Neither has an extension that is homogenous or natural enough to be used in developing an explanatory theory. Egan appears to differ from Schwitzgebel in this respect.

However, it is noteworthy that Egan does not characterize bimagination any more concretely than by saying that it has some of the dispositional features of belief and some of the dispositional features of imagination. He does not describe how the state would fit into a psychological theory, how it would interact with other mental states, and so on. Actually, it's interesting how little discussion on imagination there is in Egan's paper. There is a tension: fragmented belief is allegedly brought up to illustrate how bimagination is conceptually possible, but the account of fragmented belief seems *itself* rather well-equipped to handle the more bizarre features of delusions. It's not clear that bimagination is meant to be anything more than a type of fragmented belief. Bimaginings are said to be like imaginings in that both are circumscribed, but imagination doesn't play any greater role in the theory. Egan does not posit any sort of casual link between imagination and bimagination, for instance. This makes bimagination look a lot less like a new sort of mental category that could be used in building a psychological theory, and a lot more like a catch-all category like in-between belief. In fact, Egan has indicated to me (personal communication) that he does not think that bimagination is a unified kind. Each bimagining will come from different origins and will have different effects; no systematized theory is possible. Thus, Egan shares more in common with Schwitzgebel than is apparent from his paper.

On both views, delusions exist somewhere off the shores of paradigmatic belief.

Do Egan and Schwitzgebel offer competitor theories to the standard doxasticist account that could be used in an explanatory argument?

I maintain that they do not. To label a state a bimagination, a fragmented belief, or an in-between belief, is simply to say that it behaves like a paradigmatic belief in some respects but not others. The label will be of no help in developing a theory about *why* the state differs from a paradigmatic belief, or how it functions in a person's cognitive economy, so they cannot do the explanatory work that would be required to develop an explanatory argument for non-doxasticism. Calling a mental state a bimagination or an in-between belief offers no explanatory or predictive power over a doxasticist theory.

In fact, given that the accounts don't generate any predictions or explanations that are unavailable to the doxasticist, we can even question whether Schwitzgebel and Egan genuinely offer competitor theories. They agree with the doxasticist about the functional profile of delusions: delusions are like paradigmatic beliefs in some respects but not others. So, what exactly do they disagree with the doxasticist about? Is there even a real dispute? The only dispute appears to be whether bimaginations and in-between beliefs are non-doxastic states or whether are they are types of belief. This, I think, is a mere terminological issue. One can use the word 'belief' narrowly, to only cover prototypical beliefs that don't deviate from their paradigmatic role. Or, one can use 'belief' more broadly, to cover the less paradigmatic cases.

At this point, I digress slightly into an investigation into the nature of terminological disputes. This is needed to support my claim that Egan and Schwitzgebel offer terminological variants of doxasticism, but the digression will have supplementary benefits. Disputes over merely verbal issues dot this particular philosophical landscape like land mines. It's important for us to know how to identify and defuse them.

4.2.3 Terminological Disputes

It is easy to feel that there is nothing of any substance to non-doxasticism. There is a threat that the debate might turn on just language rather than facts about the mind. Some questions of the form "Is an X a Y?" are substantive, but others are just

terminological. "Are delusions beliefs?" looks like it might be about as answerable as the questions "Is Seabiscuit the greatest athlete of all time?", "Is gum a food?", or "Is a burrito a sandwich?" It might be that the question could be settled by deciding how to use the words 'delusion' and 'belief' rather than about what delusions and beliefs *are*.[18] This is a valid worry: the dispute between the doxasticist and the non-doxasticist should not be about words. It should genuinely be about the psychological organization of the deluded subject. Duty falls upon the non-doxasticist to explain why there is more at issue here than "mere semantics." Not all non-doxasticists succeed at doing so. Some positive proposals really aren't much more than traditional doxasticist theories gussied up in non-doxasticist language. Any debates between them and orthodoxy are merely apparent.

It would be good to know exactly when this charge is apt. How can one tell whether a debate is merely terminological or not? This is not a new question. Figures such as Carnap, Wittgenstein, and Ryle, have accused metaphysicians of treating questions of ontology with undue seriousness, when they should be dissolving the questions by paying attention to the language used to frame the debate. It is also not a settled question. There has lately been a newfound surge of interest in attempts to distinguish terminological from non-terminological debates, but no uncontentious or obviously successful criterion has emerged.[19]

Some questions are clearly non-terminological. Scientists settled the question "Is water H_2O?" by doing chemistry, not linguistics. Investigations into the atomic make-up of water did not crucially hinge on a decision about the semantics of the term 'water'. Other questions, including many discussed by metaphysicians (e.g. "Are there mereological sums?"), are ones where it is arguable whether the disagreement is terminological or not. Then, there are questions in which disputants are

[18] Consider Currie and Ravenscroft's comment: "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 the beliefs and beliefs of any other kind, then the correct response is to say that he is talking about something other than belief" (2002, p.176).

[19] See *Metametaphysics* (Chalmers et al. 2009) for a recent collection of papers largely devoted to the issue.

clearly talking past one another, as in "is gum a food?"[20] It's fun to playfully bicker this question in a kind of parody of standard conceptual analysis, proposing various necessary and sufficient conditions of gum and of food, and responding with counterexamples. But we shouldn't take ourselves to be having a substantive dispute about the metaphysics of gum. A natural reason to think there is no dispute here is that 'food' and 'gum', being natural language terms, have imprecise edges and are soaked through with indeterminacy. Some precisifications of the reference of 'food' will include gum, and other precisifications will not. Once we clarify our language, any debates over whether gum is a food should dissolve. Similarly, debates over the question "Is a cucumber a fruit?" should be resolved once we ask ourselves whether the word 'fruit' is being used in a botanical or a culinary sense.[21]

Terminological debates appear to arise when interlocutors use terms that are vague or lexically ambiguous, or when they are speaking different idiolects, or when they are using context-dependent language and there are two contexts in play. David Chalmers offers the following diagnostic test: check whether the dispute disappears after two different senses of the problematic term are distinguished (Chalmers 2009, p.88). This test might *prima facie* suggest that "Are delusions beliefs?" is merely terminological. Doxasticists apply the word 'belief' to a set of psychological states that includes delusions, whereas non-doxasticists restrict the word to a smaller set of psychological states.

However, there is a problem with the test: only the most unsophisticated use theories of meaning will hold that two words have different meanings simply in virtue of being used differently. Most theories of meaning will deny that two people who use their words differently necessarily mean different things by their words. It's not

---

[20]Hawthorne (2009) offers a list of examples of clearly terminological disagreements and a list of others that are clearly not terminological.

[21]Actually, disputes like these are sometimes taken seriously in court. For example, according to a 2006 ruling by a Massachusetts judge, a burrito is not a sandwich. A shopping complex contractually offered a Panera sandwich shop exclusive rights to sell sandwiches in their area; soon afterward, the shopping complex allowed a Qdoba burrito shop to rent space. Courts are sometimes called upon to adjudicate questions like these when the indeterminacy of language in laws and contracts leads to indeterminacy of legal and contractual norms. Questions about how judges actually do decide to precisify the semantics of legal documents, and how they *ought* to decide to do so, are of obvious interest to philosophy of law.

quite *that* easy to claim that the doxasticist and non-doxasticist are talking past one another. Chalmers's test invites us to distinguish different senses of a term, but it is not easy to know when different senses are in play. There are two different sorts of pressures–two sorts of externalism about meaning—that increase the likelihood that conversationalists will converge on the same meanings, even though they use their words differently. Firstly, as Putnam and Burge brought to attention, social factors are relevant in determining the semantics of public language terms. David Manley points out that there are many apparently verbal disputes in which people seem to mean different things by their words, but because a word is "a shared commodity whose meaning is settled by community-wide dispositions," the one who is using the term in a more unorthodox or deviant way will simply turn out to be wrong (2009, p.10). Secondly, many think that the *naturalness* of a property affects its eligibility to act as a referent of our words (Weatherson 2003, Lewis 1983, 1984). This too bears on whether a debate is merely terminological or not. Ted Sider (2009), for instance, has appealed to naturalness to defend various metaphysical debates. Given these problems, it is likely that whatever sort of epistemic deficiency is at play in terminological debates—the feature of conversations that causes us to "speak past one another"—does not come about because speakers *mean* different things by their terms.[22]

The failure of philosophers to give a precise account of terminological disputes does not mean that we are entirely unable to diagnose disputes. I propose an heuristic that does not involve having to scrutinize the meanings of our terms, also offered by Chalmers: 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

[22]Eli Hirsch thinks this. He grants that two speakers who mean the same thing in public language might still be involved in a terminological dispute and speak past one another "in their own language." According to Hirsch, X's own language is determined counterfactually: it is "that language that would belong to an imagined community typical members of which exhibit linguistic behavior that is relevantly similar to X's" (see Manley (2009, p.11)). This characterization of a language effectively gets rid of the externalist semantic constraints that are due to the public nature of language. It doesn't straightforwardly affect externalist constraints on meaning due to naturalness.

stance about whether these redescriptions are legitimate "senses" of the now-taboo word or are related to its semantic content in any particular way.

The test just described cannot give a definition of terminological disputes. It is a diagnostic that is heuristic at best, and it is not foolproof. Firstly, it will surely let in false positives and incorrectly characterize some genuine debates as terminological. There very well might be debates in which the problematic language is *primitive* and cannot be replaced with paraphrase and redescription.[23] Secondly, the test will surely let in false negatives, incorrectly ruling various terminological debates to be substantive. If a synonym of the taboo word is available, one could just replace the taboo word with it. Barring cognates might help, but picking out words that count as cognates will prove problematic. Moreover, one could introduce a new word into the language and stipulate that it is a synonym of the taboo word. (For example: "Let's invent a new word, 'schmood', and stipulate that it is synonymous with 'food'. Is gum schmood?")[24] Luckily, perfect synonymy in natural language is rare (if it exists at all). If we bar the introduction of new vocabulary, most attempts to replace a word with a description should precisify the terms used, either dissolving the debate or making clear what issues of substance are being disputed.

4.2.3.i Are *Delusions* Beliefs?

Now that we have a diagnostic test in hand, let's see how it can be put to use. There are two potentially problematic terms in the assertion "delusions are beliefs."[25] Let's first focus on potential indeterminacies in the word 'delusion'. It is not an especially

---

[23] For instance, many of the metaphysical debates suspected to be terminological involve existence assertions, such as "there are mereological sums." It does not appear that anyone here is using the term 'mereological sum' in a different way; the problem, if there is one, appears to come from semantic variance in the quantifier. This is a failing of the new diagnostic test. It might not be possible to conduct any sort of metaphysical discussion at all if we are robbed of terms such as 'there are', but not all ontological debates involving existence assertions are bankrupt. Similarly, disagreement among ethicists cannot be ruled terminological simply because the disagreement cannot be conducted without normative vocabulary such as 'ought'. See Chalmers (2009).

[24] Synonyms are not the only problematic stipulations; any definition mentioning the banned term will be problematic. Antonyms, for example. There is little progress to be made in debating whether gum is a "nonfood."

[25] Specifically, 'delusions' and 'beliefs'. I presume that 'are' is benign.

problematic word—aside for some fuzziness at the fringes, participants in the doxasticism debate largely agree on which sorts of individuals are delusional. However, not everyone agrees about which mental state of the deluded individual is the delusion. Some non-doxasticists hold that delusions are rich and involve multiple psychological states, including *experiences* (Spitzer 1990, Parnas and Sass 2001, Ratcliffe 2008). These experiences can *lead* to doxastic states with the delusional content, or a delusion can be a complex that *comprises* the doxastic states as well as the experiences, but the doxastic states themselves are not the delusions.[26]

This looks like it might be an entirely semantic dispute over the proper referent of the word 'delusion'. You and I might have the exact same theory about the etiology of Capgras syndrome. We might agree that in patients suffering from Capgras, a disordered experience causes the patient to have false beliefs about their loved ones.[27] However, we disagree on whether the *experience* is the delusion or the *belief* is the delusion. The key point here is that we agree on the functional architecture of the delusional individual, and we can't begin to articulate our disagreement without using the word 'delusion.' That's reason to think that there's no debate here except over the term.

Here is George Graham (2011) describing the view he has developed with Lynn Stephens:

> I do not wish to classify delusions as beliefs (and I doubt whether thinking of delusions as beliefs helps all patients). This is because I think of delusions as messy, compound, and complex psychological states or attitudes (thoughts, feelings, and so on), defined more by how persons mismanage their content and fail to prudently act in terms of them, than by qualifying as beliefs (see Stephens and Graham 2004, 2007, Graham 2010).

One could agree that delusions involve messy complexes of multiple mental states that include beliefs, but disagree about whether the term 'delusion' should affix to the complex or to a belief within the complex. The phrases 'classify', 'thinking of',

[26] I should note that this is not necessarily an argument that the authors cited would themselves endorse; I am extracting it from their claims about the richness of delusion. These authors (and Graham, below) have other reasons for championing non-doxasticism that are not simply terminological.

[27] This is Maher's (1974) explanation.

and 'qualifying as' all suggest that what is at issue is how we should best stipulate the meaning of a term. This is surely an important debate to have; there are important pragmatic reasons for psychiatrists and therapists to prefer a clinical term to have one meaning rather than another (see McHugh and Slavney (1998, p.64–5)). But this is to say that some terminological debates are important, not that this is not a terminological debate.

The formulation of non-doxasticism that I put forward in the first chapter avoids any problems on this front. I proposed that non-doxasticism about delusions is the conjunction of two theses: a conceptual thesis that that one can have the delusion that *p* without believing that *p*, and an empirical thesis that delusional patients in fact do not typically believe the contents of their delusions. One can think that the word 'delusion' should properly affix to a "messy, compound, and complex" set of psychological states, and still ask whether there must necessarily be a belief at its core. Non-doxasticists can avoid using the word 'delusion' entirely by ostending to delusional subjects and denying that those subjects believe the wild assertions that they affirm.

4.2.3.ii Are Delusions *Beliefs*?

The word 'belief' is a more worrisome source of indeterminacy than the word 'delusion'. As Mallon and Stich (2000) argue, many apparent debates about belief reduce to semantic disputes about the extension of 'belief'. Suppose you and I are debating whether trucks are cars. You think the word 'car' refers to a class of vehicles that includes the Chevy Silverado (a type of truck), and I think it does not. We both agree that trucks have pickup beds or cargo holds. You are committed to the idea that cars do not have pickup beds or cargo holds; I think they can have them, but not necessarily.

It is not easy (if it's at all possible) to translate this debate into a language lacking the word 'car'.

Now let's draw out an analogy. You think the word 'belief' refers to a class of mental states that includes the very peculiar ones seen in schizophrenics, and I think it

does not. We both agree on the patterns of inference the mental state is disposed to engage in, and we agree on its functional profile, but you differ in affixing 'belief' to states with this functional profile. This mirrors the car/truck example. It appears extremely difficult to have this debate in a language lacking the term 'belief'.

A potential problem lurks here. We previously worried that one way to circumvent the test would be to introduce a synonym of the banned word. Many non-doxasticists *do* introduce new vocabulary (such as 'bimagination') in order to articulate their positive proposals. Their doing so cloaks the similarity between them and doxasticists.

Let me sketch out a purposefully ill-conceived non-doxasticist theory. I hereby posit a new mental state called a *schmelief*. A schmelief has many of the features of a belief, but it is poorly integrated with other beliefs, it isn't sensitive to evidence, it controls verbal behavior but otherwise isn't used in practical reasoning, etc. Delusions are not beliefs; they are schmeliefs.

This "theory" contributes nothing: it just describes the functional role that delusions have (in terms of belief, no less) and affixes a silly name to it. It's certainly not a theory with any additional explanatory bite; it's rather a piece of linguistic stipulation. There's no reason to think that a schmelief is anything other than a kind of belief that we've needlessly rechristened. If someone were adamant that only rational beliefs are entitled to the title 'beliefs', then that person might prefer to call delusions 'schmeliefs', but this would simply be a terminological move. Arguing about whether delusions are schmeliefs is like arguing about whether gum is schmood.

Many of the non-doxasticist proposals out there do no better than schmelief theory. Usually, non-doxasticists who are open to this sort of charge will argue for claims that are genuinely contentious and non terminological, but those claims are ones that doxasticists could happily hold, and the debate they are having should properly be considered a civil war within doxasticism. Bortolotti rightly complains that certain non-doxasticist proposals are so imprecisely stated that they do not give her much of a grasp on how they are different from doxasticist accounts, and she wonders whether they really say anything different at all (Bortolotti 2009b, p.169). She is,

in effect, charging non-doxasticists with positing schmeliefs.

4.2.4 Criticism of Egan and Schwitzgebel

Let's ban the terms 'bimagination' and 'in-between belief'. What can we say about Egan's and Schwitzgebel's proposals? The most bare-bones description of their views is that delusions have some of the dispositional features of prototypical belief, but not all. However, this is not likely a claim that a doxasticist will argue with! Of course delusions do not act like prototypical beliefs; this is why we consider them pathological. A doxasticist would want to *call* bimagination a type of belief, and Egan might not, but it looks as if this is simply a terminological dispute: Egan and Schwitzgebel and the doxasticist can agree on all the underlying dispositional facts. If the lesson to take away from Egan's paper is that we should think of delusions simply as fragmented beliefs, this is simply to agree with Bortolotti that delusions are irrational beliefs: the description of fragmented belief that Egan offers simply *is* a description of inconsistent belief. What could inconsistent belief be if not just fragmented belief? To believe p while also inconsistently believing not-p is just to manifest a belief that p at some times and in some respects, and a belief that not-p at other times and in other respects. Bimaginations and in-between beliefs are schmeliefs.

Bortolotti takes the debate between doxasticism and theories of in-between belief to be substantive, and not merely verbal. She briefly considers the notion that 'belief' is vague and that delusions might be a mental state that are only somewhat like beliefs, but she rejects this on the grounds that it complicates policy applications and ethical judgments (2009b, p.20). Schwitzgebel (forthcoming) rightly retorts that these issues probably *should* be complicated. This debate very much has the feel of a terminological debate. Bortolotti might be right: it might matter what the we choose to affix the word 'belief' to for policy purposes, and it may be that we should choose something simple. It also might matter what we choose to affix the word 'delusion' to for diagnostic purposes... or what we choose to affix the word 'vehicle' to for the purposes of drafting traffic laws. This does not mean these debates are not terminological.

To some extent, it is unfair to lump Schwitzgebel and Egan in with the doxasticists. Their accounts *do* militate against doxasticists who adamantly try to explain the weirdness of people's behaviors in terms of normal, prototypical beliefs. For example, a doxasticist might be wedded to a psychodynamic perspective that attempts to explain psychoses in terms of covert beliefs and desires; a doxasticist might think that delusional subjects are rational, but some sort of performance error is corrupting just the outward expression of the mental state. Egan and Schwitzgebel can be taken as cautioning theorists against a blinkered search only for explanations in terms of rational belief-desire psychology.[28] However, it is not clear that many doxasticists merit this reprimand.

Even if their current accounts are doxasticist, Egan and Schwitzgebel's accounts could serve as the inspiration for a genuinely non-doxasticist theory. If Egan were to develop a full theory about how bimaginations are formed, and a systematic account of its relations with the rest of the cognitive economy were developed, then it could be a genuine alternative to doxasticism. To describe these relations would be to describe the contours of a new functional role. If the functional role were one that doxasticists do not currently posit and could not easily be assimilated to a form of fractured belief, then the theory would make a claim beyond a mere terminological suggestion. Consider how difficult it would be to assimilate an attitude such as *desire* or *anger* to non-prototypical belief. A genuine non-doxasticist proposal should offer something similarly difficult to assimilate. Egan's focus on imagination is suggestive and promising: a consistent story about the imaginative component of delusions, and a description of how delusions interact with imaginings, would give the theory some heft. However, as previously mentioned, Egan is pessimistic about these prospects.

---

[28] Egan mentions that the standard "boxological" metaphor might be misinterpreted as implying that all mental representations must fit nicely within boxes.

4.3 Framework Propositions

4.3.1 Campbell's Framework Proposition Account

One of the more popular non-doxasticist accounts of delusions claims that delusions are Wittgensteinian *framework propositions*.[29] The suggestion was first made by John Campbell in his "Rationality, Meaning, and the Analysis of Delusion" (2001), and the paper remains the canonical text. Campbell was inspired by some of the later Wittgenstein's writings on skepticism. In *On Certainty* (1975), Wittgenstein opined that when confronted with a doggedly persistent philosophical skeptic, there comes a point where we reach beliefs that we hold with conviction—epistemic bedrock—for which we do not feel pressured to provide justification. These "framework propositions" are beyond question:

> "If someone wanted to arouse doubts in me and spoke like this: here your memory is deceiving you, there you've been taken in, there again you have not been thorough enough in satisfying yourself, etc., and if I did not allow myself to be shaken but kept to my certainty—then my doing so cannot be wrong.
>
> The queer thing is that even though I find it quite correct for someone to say "Rubbish!" and so brush aside the attempt to confuse him with doubts at bedrock—nevertheless I hold it to be incorrect if he seeks to defend himself." (Wittgenstein 1975, p.497–8)

The analogy with delusions is fairly clear. Examples of framework propositions given by Wittgenstein include "there are a lot of objects in the world" and "the world has existed for quite a long time." Campbell holds that the delusional patient has pathologically assigned this status to bizarre propositions such as "I am dead."[30]

---

[29] For papers supporting a framework proposition account, see Campbell (2001), Klee (2004b,a), Rhodes and Gipps (2008, 2011), and Welz (forthcoming). For arguments in opposition, see Bayne and Pacherie (2004), Thornton (2008), Bortolotti (2011a), and Bortolotti (2009b, p.187–197). Wright (2004) has probably done the most work attempting to make framework propositions philosophically respectable.

[30] Amusingly, Wittgenstein (1975) also gives the following example: "If we are thinking within our system, then it is certain that no one has ever been on the moon. Not merely is nothing of the sort seriously reported to us by reasonable people, but the whole system of our physics forbids us to believe it. For this demands answers to the questions 'How did he overcomes the force of gravity?' 'How could he live without an atmosphere?' and a thousand others which could not be answered" (p.108). With unfortunate timing, this text was first published (posthumously) in the same year as the moon landing. The mention of "thinking within our system" demonstrates the difference between framework propositions

Campbell stresses three features of framework propositions that are central his account of delusions. Firstly, because a person feels no epistemic pressure to provide reasons for their framework propositions, the framework propositions are immutable and not subject to ordinary rational revision. Just as delusional subjects do not feel pressure to revise their delusions at the urgings of medical doctors, so too do we not feel pressure to revise very central bedrock commitments at urgings of doctors of philosophy.

Secondly, Campbell presents an etiological account of delusions. He holds that because framework propositions are outside the realm of rational justification, delusions are accordingly not propositions that we have rationally reasoned our way into. They have not been formed for reasons. Instead, they are changes in belief caused by direct "organic malfunction." (Think of a person who is bumped on the head and forms a new belief simply in virtue of the neurological damage suffered.) This is to be contrasted with accounts positing that delusions are the product of broadly rational inferences from unusual experiences. Campbell calls these latter sorts of accounts "empiricist" or "bottom-up," and he calls his own account a "rationalist" or "top-down" account. These terms have become widely adopted and are now common coin in all sorts of debates about delusions, whether the debates are about framework propositions or not.[31] On a rationalist theory, if a delusional patient has odd experiences, this is the result of the top-down influence from the framework belief.

Thirdly, framework propositions allegedly have a special kind of semantic status. Campbell writes,"Wittgenstein's notion of a framework proposition was never

---

and the sorts of foundational beliefs that Descartes had in mind; clearly, Wittgenstein means for beliefs to only count as framework propositions relative to particular contexts or standards. Also, note that although Wittgenstein claims that it is a mistake to supply justifications for framework propositions, he curiously supplies justifications for one in this passage. (To be fair, *On Certainty* is composed of unfinished fragments published posthumously, and it is probably unfair to demand much consistency of it.)

[31] Unfortunately, they are not always used consistently. Campbell's explicit definitions of empiricist and rationalist accounts do not seem to classify "two-factor" etiological accounts, in which a delusion is formed in response to an unusual experience, but the response is irrational. However, he does classify two-factor theorists such as Andrew Young as empiricists, suggesting that the crucial feature of empiricist accounts is that delusions are formed in response to experiences (whether rationally or irrationally formed). These are more useful definitions of 'empiricist' and 'rationalist' anyhow, as we already have the terms 'one-factor' and 'two-factor' to respectively designate theories that involve rational and irrational responses to experiences. I'll continue to use the terms in this way.

worked out in great detail. But it is certainly part of the picture here that a change in framework principles would bring with it a change in the meanings of the terms used" (Campbell 2001, p.98). Campbell is motivated by arguments of Quine's and Davidson's that the significance of our terms is fixed by the use we make of them in rational reasoning. A particularly important form of reasoning for fixing the meaning of our terms is a *verification procedure*: we must be able to know what sorts of situations would render a proposition true and which would render it false in order to legitimately grasp it. In fact, the theory of meaning that Campbell favors teeters toward verificationism. Because framework propositions are treated as background assumptions, they are "not themselves, in any ordinary way, subject to empirical scrutiny" (p.96). And since the Capgras patient, for example, "does not use [the canonical rational] way of checking who it is that is before him, he seems to have lost his grip on the meaning of the word" (p.91). In fact, Campbell claims that the delusional subject speaks an entirely different language than the rest of us: "The situation is rather as Kuhn (1970) described the relations between the terms used in scientific theory before and after a revolutionary change" (p.98).

It's important to realize that the three pillars of Campbell's characterization of framework propositions—immunity to revision, top-down etiology, and semantic deviation—are all dissociable. They needn't necessarily go together. For instance, take the first and the second pillars. Bayne and Pacherie note that Campbell and Wittgenstein offer "there are many tables and chairs in this room" as a framework proposition, but we surely acquire that proposition from experience and not from a top-down organic defect. And in the other direction: if it is possible for us to acquire beliefs from a bump on the head without reasoning our way to them, it is not at all obvious that reason could not kick in at a later time to have us revise them. (At the very least, if we can't revise beliefs that arise from a bump on the head, this would be an empirical truth and not the conceptual truth that Campbell presents.) As far as the third pillar goes, the first two pillars do not on their own imply that the delusional subject means something different with his words. Campbell requires his verificationist theory of meaning to get this conclusion. Nonetheless, it is common for

advocates of framework propositions to subscribe to all three pillars (e.g. see Rhodes and Gipps (2008)).

4.3.2 Criticism of Campbell

Is there any reason to think that the framework proposition theory offers a genuine positive proposal and is not just a terminological variant of doxasticism? Not especially. The most plausible version of the theory is easily statable within a doxasticist framework. A *sui generis* class of framework propositions is not of any explanatory value not available to the doxasticist.

We might ask why Campbell's proposal is often taken to be non-doxasticist. (Bortolotti targets it in her book, for example.) There are two general ways of creating a non-doxasticist positive proposal: by altering the attitude or the content. One can offer up an attitude that is distinct from belief, or one can argue that delusions are beliefs but the content is not what it appears to be. Campbell's proposal can be read in either way. One might think the attitude we bear toward framework propositions is not, properly speaking, a belief.[32] One might additionally be moved by Campbell's semantic theory and conclude that the delusion with the apparent content that *p* is not about *p* at all. However, the former is not a substantive claim and the latter incorporates a semantic theory that is not very plausible at all.

Let's first get rid of the latter, implausible claim: the claim that the delusional patient's words have an entirely different meaning. Here is a strange implication this would have, were it true: if framework propositions are relieved of needing "empirical scrutiny," and it is this feature that causes their meanings to be nonstandard, then we would expect that any framework proposition we hold has nonstandard meanings. I can use the word 'table' in a sentence that does not express a framework

[32] It's not obvious whether Campbell's proposal is meant to posit a new attitude or not—the language in this literature is often tricky. Just as the word 'belief' can refer to the content of our mental states rather than our mental states themselves (as in "I disbelieve all your beliefs"), and the term 'framework proposition' is surely meant to similarly refer to a content. Thus, it's not transparent what sort of attitude Campbell thinks we bear toward the framework propositions. He does occasionally call them beliefs, but elsewhere intimates they are not beliefs.

proposition ("Ikea is selling tables for 20 percent off tomorrow"); does the word 'table' in that sentence have a different meaning than in the sentence "there are many tables in the world?" Campbell's claim that the delusional patient means something else with his words does not in any way emerge from his commitment to delusions being framework propositions. Rather, it has to do with his theory of meaning, along with his observation that delusional subjects make odd assertions. However, except in extreme cases, delusional subjects use words broadly in the same way that we do. The Capgras patient knows what spouses are, and we can converse with him about them. In the previous chapter, we rejected the Berrios theory that delusional patients are uttering empty speech acts because we are able to have conversations with them about their delusions. The same evidence applies here: we appear to be able to communicate with delusional subjects. Any theory of meaning that claims that we fail to communicate with one another simply because we use our words differently in some way is much too strict.

The other two features of framework propositions make up a theory of delusions that is at least halfway plausible. The theory is, however, just doxasticism in disguise. Consider the following quotation of Rhodes and Gipps:

> It is true that at various junctures in *On Certainty* Wittgenstein talks of certain propositions—the 'hinge propositions'—as lying at the certain foundations of the language game. Nevertheless, these propositions and the beliefs they seem to express have so little in common with what we should everyday call a proposition and a belief—what with their (logically) not being able to be proved, evidenced, described as empirically true or false, justified, and so on—that they barely qualify as such. (Rhodes and Gipps 2011, p.94)

These features might be foreign to the beliefs we standardly talk about in everyday speech, but are they really unavailable to the doxasticist's conception of belief? Immunity from revision is not difficult feature for a doxasticist to handle. Everyone has beliefs that they are unlikely to revise. Quine called these "central beliefs": they are at the center of our web of belief and deeply ingratiated into our theory of the world. They are for all practical purposes immune to revision and never very much threatened by recalcitrant observations: we will always make adjustments in our theory

elsewhere in our web of belief. Whenever an argument tells us that one of our central beliefs is false, we will opt to reject one of the premises instead. It is also possible to put this explanation in terms of credences. There are some propositions in which we have a very high credence, and our subjective conditional probabilities are such that that whatever evidence we are likely to encounter will not substantially lower our credence.

Campbell might complain that framework propositions are importantly different from central beliefs. Central beliefs are ones that we can offer very strong justification for holding: they are firmly enmeshed in our web of belief and receive holistic support from all sorts of other propositions we believe. Framework propositions, on the other hand, are supposed to be propositions that do not receive any sort of support from anything else (and are not treated as if they require support). However, this vitiates the claim that delusions are framework propositions: delusional subjects do not feel that they need provide *no* justification. They feel pressure to explain what they're saying, and they typically confabulate responses. If you ask me how I know that my arm is my own, I might gun you down with an incredulous stare. Delusional subjects *do* try to give you reasons. ("It's my mother's arm.") Additionally, it is not a problem for the doxasticist to posit a rationalist etiology. A doxasticist can happily say that a stroke directly alters a patient's credence in the proposition that the woman who is before him is not his wife.

What should we say about this view, doxasticist as it is? It's plausible, and I think its simplicity recommends it. However, it doesn't account for the data that most strongly pushes one toward doxasticism. The *effects* of delusions are not as expected: delusional subjects don't appropriately act on their beliefs. Moreover, they seem to experience "double-bookkeeping." These features are the mysterious pieces of evidence that many accounts fail to get right, and the "central belief" account fails to address them. The only portion of Campbell's theory that attempted to explain *these* data was the portion that claimed that the *content* of the delusions became mangled, given their status as framework propositions.

4.4 Summary

A number of philosophers have attempted to offer non-doxasticist accounts of delusion that would serve in explanatory arguments. These accounts tend to fail for one of two reasons. Firstly, they can fail to be any more explanatory than standard doxasticism, because they amount to simply a redescription of a doxasticist account using different vocabulary. Egan's, Schwitzgebel's, and Campbell's theories exhibit this failing. Doxasticism admits that delusions are unlike paradigmatic beliefs—it does little good to simply give the non-paradigmatic states a new name. True non-doxasticist explanations of the oddities of delusion, rather than a simple restatement of the phenomena, would need to say that the oddities are explained by delusions being *X*s because we have a theory of *X*s that can account for those features. Secondly, attempted non-doxasticist accounts can fail because they do not accurately predict the features that delusions exhibit. Currie's theory exhibits this failing. We would not expect metarepresentational beliefs to act as delusions act.

In the next chapter, I make steps toward developing a non-doxasticist account—one that is not a mere terminological variant of a doxasticist account—by assuaging the worries of those who would quail against novel belief-like attitudes. I diagnose why doxasticists might be tempted to reject novel belief-like attitudes, and argue that a radical account that posits acceptances does not carry the burden of being intolerably unintuitive and revisionary by showing that our understanding of belief is shakier than it might at first appear,

Chapter 5
Concepts of Belief

5.1 Positing Attitudes

The radical non-doxasticist is in the business of *positing attitudes*. This is a revisionary project. Psychology is concerned with determining the way the mind works, not just determining the way the man on the street thinks his mind works, and so the psychologist is free to develop previously unconsidered sorts of mental architectures.

A success story comes from the field of memory research. Cognitive science has revealed that the ordinary concept of memory denotes a number of very different mental processes, such as working memory, semantic memory, and episodic memory. The differences between these various faculties are not obvious on the surface; psychologists had to discover them. Recent work in cognitive neuroscience is also pulling apart the ordinary notion of imagination (Addis et al. 2009). Psychopathology is an especially valuable field of study when looking for evidence that our folk concepts of the mental can be fragmented. Psychopathologies often present cases of selective impairment, where one process or faculty is disturbed and another is not. For example, individuals with anterograde amnesia can act on short-term memories but are unable to fix long-term memories, indicating that there multiple mechanisms and faculties at work in memory.

Given the successes found in memory and imagination, we should perhaps expect that ordinary folk concepts are what Block (1995) calls "mongrel concepts." A mongrel concept is a concept that picks out various dissimilar states. It would be surprising if ordinary intentional idioms were already honed to perfect precision, unable to be improved upon. If the folk concept of BELIEF is a mongrel concept,

we would expect to find evidence of this hypothesis by considering pathologies of belief, and delusions look like a perfect phenomenon upon which to direct out attention. Delusions only partially behave like prototypical beliefs—patients tend to say things without acting on them—and this could be because belief is a mongrel concept that denotes multiple types of mental states produced by different sorts of mental faculties, only some of which are impaired in delusional subjects. Murphy puts the point nicely: "[p]erhaps the very diverse causes of belief that folk thought recognizes is evidence that belief is not a natural kind, and that a mature psychology should recognize a number of different sorts of intentional state with different relations to each other and to behavior" (Murphy 2012).

The account that I propose posits an attitude that is not typically recognized by philosophers: a delusion is an *acceptance*, which is like a belief in many respects, but is distinct from belief. One can accept that one's wife has been replaced with an imposter without truly believing it, and this is the situation of the Capgras patient. However, many have urged caution about adopting too revisionary of a psychology; conservativism of this sort can result in the thought that radical non-doxasticist account must be getting something wrong. I have found that many philosophers recoil at radical non-doxasticism in this way, and their worries should be addressed. An opponent might say,

> Look, delusional subjects may not always act on their delusions or reason about them appropriately, but they do confidently assert them. Delusions are very similar to other sorts of irrational beliefs, so in order to claim that delusions aren't beliefs, you need to also maintain that all sorts of other superstitions and ideological convictions are not beliefs. That goes too far; superstitions and ideological convictions are beliefs if anything is. You must be misusing the term 'belief'. You might be able to develop a theory of a *subtype* of belief, and convince me that delusions are this sort of subtype. This is what happened in memory research. We didn't discover that what we thought were memories turned out not be memories. We discovered that there are *different types of memory*. Scientific progress will be made by investigating belief, not casting it aside. *Call* your new state a non-belief, if you must. Call it whatever you want. That doesn't make it a non-belief.

The goal of this chapter is to defray these sorts of concerns. One possible response would be to flat-footedly maintain that psychology can be as revisionary as it needs to be, and while I have sympathies with this maneuver, I do not think it necessary in this case. Rather, it is possible to show that varieties of non-doxasticism are not as revisionary as they might at first appear. The hypothetical opponent above charges me with misusing the term 'belief', but the standards for correct use of the term are, I plan to show, hardly well-established or easy to articulate. The term is used unclearly and inconsistently in technical philosophy papers and in folk discourse alike. Categorizing delusions—or even religious convictions or superstitions—as non-beliefs is not an unacceptably revisionary affront to either folk psychology or to academic philosophy, because in neither case is there a very principled practice of typing beliefs. Doxasticism is partly fueled by a sentiment of conservativism, and this sentiment can be undermined by examining just poorly understood and how open to reinterpretation the attitude of belief actually is.

Over the course of this chapter, I'll pursue three goals. Firstly, I will describe the typical philosophical conception of belief, and show that it is not as clear-cut or well-defined a theoretical posit as one might think. Any characterization of belief (that is not obviously faulty) is impressionistic enough to provide ample room for thinking that there are various sorts of belief-like attitudes, and for thinking that the traditionally philosophical notion of belief might play a smaller role in our cognitive lives than is often presumed. Secondly, I'll describe the role that BELIEF plays in folk psychology and the role that 'belief' plays in the folk vernacular, and establish that there is a mismatch between the philosophical notion of belief and the folk notion. Folk psychology is not a great place for the defender of the traditional philosophical notion to turn to; the folk notion fits the philosophical notion imperfectly, and equally well supports the contention that there is a split between belief and acceptance. Finally, I'll explain why a theory that posits acceptance is not a mere terminological variant of doxasticism.

Before moving on to close examinations of the philosophical and the folk notions of belief, I begin with a brief comparison of the two.

5.2 The Lush and the Sparse

Belief-desire psychology holds that we act on our beliefs to fulfill our desires. Fodor opens his *Psychosemantics* (1987) with an analysis of a scene from *A Midsummer Night's Dream* in the terms of belief-desire psychology. Here is a sample segment:

> Hermia believes (correctly) that if *x* wants that *P*, and *x* believes that not-*P* unless *Q*, and *x* believes that *x* can bring it about that *Q*, then (ceteris paribus) *x* tries to bring it about that *Q*. Moreover, Hermia believes (again correctly) that, by and large, people succeed in bringing about what they try to bring about. *So:* Knowing and believing all this, Hermia infers that perhaps Demetrius has killed Lysander. (Fodor 1987, p.2)

Fodor presents this tale in order to make a point about the sorts of inferences that could be implicitly going on the deep recesses of the reader's non-conscious mind. It may well serve as an accurate description of that. Still, it certainly sounds artificial and artless, as do many explanations of intentional action that are cashed out in belief-desire terms. One reason for this might be that there is something ungraceful in using the word 'belief' so consistently. Belief and desire have a privileged place in discussions of belief-desire psychology (hence the name), and non-philosophers can be taken aback by the prominence of these two particular mental states. The words 'belief' and 'desire' do not figure in everyday talk as much as a great many other propositional attitudes.[1]

The tools that novelists and playwrights have at their disposal are many and varied. Shakespeare is able to subtly distinguish between various mental attitudes in our protagonists using a panoply of colorful expressions. There is something lumpen about trying to capture the inner psychologies of our protagonists using just a handful of propositional attitude verbs. Our stock of intentional idioms is *lush*, not *sparse*. We hope, we suspect, we consider, we accept, we maintain, we think, we feel, *et cetera*.

[1] An informal search on the *Google Books Ngram Viewer*, which searches a corpus of all books that Google has scanned, reveals that 'believe' is used about half to two-thirds as frequently as 'know', 'see', or 'think'.

Yet, there is a consistent drive in the philosophical literature for a sparse theory of mind, and in this sparse theory, belief and desire take a place of pride. "Many philosophers think that belief and desire are somehow the most basic Intentional states," Searle contends (1983, p.29), and evidence for this claim is in no small abundance. Lewis writes, "I limit my attention to these attitudes in the hopes that all others will prove to be analyzable as patterns of belief and desire, actual and potential" (Lewis 1974, p.332). Why the hope? And why those two? Epistemologists who model thought using the tools of decision theory do so using credence and utility functions, representing graded versions of belief and desire. Are there principled reasons to use only these two, or is this a hand-me-down assumption? In the 1970s, Gilbert Harman (1976) proposed that belief and desire were not sufficient to explain our abilities to engage in practical reasoning. We also needed to posit *intentions*, a class of mental attitudes distinct from desires, wishes, hopes, aims, and predictions. Davidson, to whom Harman was largely responding, was unwilling to countenance intentions in addition to beliefs and desires. Why?

Doxasticists who argue against radical accounts of delusion might simply be responsibly conservative, demanding a wealth of evidence before we amend or overturn traditional belief-desire psychology. However, I think something deeper is going on; the evidence just mentioned suggests a tendency for philosophers to think that belief-desire psychology is, and should remain, sparse. An assumption of this sort would produce a bias toward doxasticism. However, it is not clear that there is principled reason to hold onto this assumption. We should be skeptical: the drive for sparseness, and the drive to elevate belief and desire above other attitudes in conceptual and explanatory priority, both appear to fly in the face of the actual vernacular, which is lush.

5.3 'Belief' in the Mouths of Philosophers

In order to determine where this drive toward sparsity comes from, we need to examine the philosophical conception of belief. What do philosophers think that beliefs

are?[2] Typical investigations into the nature of belief compare various theories of the propositional attitudes, such as representationalism, dispositionalism, or functionalism. However, my concern is not with the metaphysics of mental states in general. My concern is with the features of belief that philosophers think distinguish it from other mental states.

It is often recognized that the philosophical use of the word 'belief' deviates from the way the word is used in folk discourse; the word is a term of art, allegedly with a technical use. The word tends to be used in two ways. 'Belief' is sometimes used as a very general umbrella term for *informational* mental states—those with a mind-to-world direction of fit. At other times, belief is characterized as a state that plays a particular role in our cognitive economy and that enters into particular relations with other mental states. Let's consider these in turn.

5.3.1 'Belief' as an Umbrella Term

'Belief' and 'desire' are often used as umbrella terms for states that have an appropriate *direction of fit*. Beliefs have mind-to-world direction of fit. It's this sort of usage, I'll argue, that fuels the philosophical drive toward sparsity—toward thinking of belief and desire as "the most basic Intentional states." (Searle 1983, p.29).

It's easiest to describe this use of 'belief' by considering the construction of a very simple robot. Suppose that you needed to program a robot to accomplish a certain goal. One strategy would be to have it always perform a certain behavioral sequence: a robotic arm on an assembly line, for instance, might need to do nothing more than repeat a series of predetermined movements. This strategy works in the case of an assembly line because we can assume an unchanging environment. Whenever the

[2] Obviously, different philosophers have different ideas about the nature of belief, so there is a danger of being too procrustean and curt in talking about a single way that philosophers use the term. There are major debates about whether beliefs can be formed voluntarily, whether we should think of belief as a categorical state or whether we should think instead of degrees of confidence, and so on. However, there do exist received standards and conventions that govern talk about belief in the pages of a philosophy journal. A paper that deviates from these standards without flagging the deviation will generate raised eyebrows, confusion, or charges of incompetence. The claims that I make are general claims about belief-talk in the establishment of analytic philosophy; I can simultaneously recognize that many—perhaps even most—philosophers have their own idiosyncratic beliefs about belief.

arm bends down, the conveyer belt will have pushed forward a new object for it to manipulate. However, this strategy will not work in a changing environment. We will need to make our robot more complex for it to register the state of the environment that it is currently inhabiting. If two different sorts of objects can appear on the conveyer belt, each demanding a different routine from our mechanical arm, perhaps we should graft on a sensor so that it can react appropriately to each. We thereby give the arm "beliefs" of a sort—informational states about what is in front of it.[3] Here are a few passages that further describe this strategy:

> [W]e have every reason to believe that [all but very simple] animals are implementing a behavioral strategy that factors the problem of how to behave into two problems—the problem of harvesting, processing, and storing information, and the problem of settling on courses of behavior depending on what information is made available by harvesting, processing, and storing information. It is, frankly, a strong point in favor of belief-desire psychology that it predicts the factorization strategy that led to the demise of behaviorism and the rise of cognitivism. (Ichikawa et al. 2012, p.4)
>
> If we think about behavior-planning systems as systems for figuring out ways to get from some start state (the way things are taken to actually be, right now) to some goal state (the way things are desired to be) we can characterize this difference in the roles of belief-type and desire-type representations this way: My behavior-planning systems look at belief-type representations in order to determine *start* states, and they look at desire-type representations in order to determine *goal* states. (Egan 2009, p.264)

The distinction between start states and goal states, or between states that represent the way the world is and the way the world should be, or between the cognitive and the conative, might be thought of as a basic distinction that must be present in any reasonably intelligent agent, and 'belief' and 'desire' (or 'belief-type' and 'desire-type') are sometimes used as general terms to cover these two types of state.[4]

This distinction is part of Searle's diagnosis for the perceived "primacy" of belief

---

[3] I use scare quotes because this use of 'belief' does not comport especially well with the vernacular. Nonetheless, the term is often used by philosophers to apply to very basic informational states employed in even very simple behavioral strategies.

[4] Bratman (1987) and others have argued that *intention* must also be posited in agents sophisticated enough to form plans of action. His belief-desire-intention model of practical reasoning has formed the basis of "BDI software models," used in programming artificially intelligences; this is a nice example of a philosopher having made a contribution to software engineering.

and desire and the drive toward sparsity. Searle notes that belief and desire are often used by philosophers very broadly, in a way that departs from ordinary English, "corresponding roughly to the great traditional categories of Cognition and Volition" (1983, p.30). Beliefs include convictions, hunches, inklings, and acceptances—any state with mind-to-world direction of fit.[5] Desires include wants, wishes, lusts, and hankerings—any state with world-to-mind direction of fit.

If belief and desire are conceived of this broadly, then one might think that all behavior can be explained in these terms. Searle considers whether all other mental states are reducible to complexes of belief and desire. An expectation that *p*, for example, might simply be a belief that *p* will be true in the future. A fear that *p* might be a belief that *p* is possible and a desire that *p* not be actual. Remorse that *p* might be a belief that *p*, a belief that I am responsible for *p*, and a desire that *not-p*. A belief, or a hope, that such reductions are possible, would explain why Lewis thought he might be able to get away with only positing beliefs and desires, and why decision theorists use only credence and utility functions when modelling behavior. (It is, in fact, hard to see where else the urge to interpret human behavior solely in terms of beliefs and desires could come from.)

However, there are three major problems with treating 'belief' and 'desire' as umbrella terms and with thinking that we have all our other intentional states in virtue of our beliefs and desires.

Firstly, as Searle argues, the attempted reductions are insufficiently precise; they are not fine-grained enough to distinguish between importantly different mental states. Being annoyed that *p*, being sad that *p*, and being sorry that *p* all imply a belief that *p* and a desire that not-*p*, but any differences between them cannot be otherwise cashed out in terms of belief and desire. Being amused that *p* implies that one believes that *p*, but no other beliefs or desires can be extracted, and there is clearly more to amusement than simply belief. Fear is not simply a combination of a belief that *p*

[5] If a state with mind-to-world direction of fit represents *p* as obtaining but *p* does not in fact obtain, then the state has *failed* in some way. States with mind-to-world direction of fit are truth-directed. Direction of fit is therefore a normative notion. (Velleman 1992).

is possible and a desire that *p* not be actual, for one can hold these beliefs and desires without experiencing fear. Fear also has, at the very least, a particular affective feel.

Secondly, not all mental states have a direction of fit. Searle disagrees with this claim. He holds that mental states cannot be reduced to collections of belief and desire, he does maintain that all intentional state attributions *imply* something about belief and desire. According to his theory of intentionality, every intentional state has conditions of satisfaction that are met when either it fits the world or the world conforms to fit it. However, he is wrong about this. Take, for example, imagination, assumption, pretense, or supposition. I can imagine that *p* without this implying anything about my beliefs or my desires. Imagination does not appear to have a "proper object" in the way that truth can be said to be the proper object of belief, the good can be said to be the proper object of desire, and the dangerous can be said to be the proper object of fear. Searle does, in fact, list "supposition" among the mental states that the very general category of 'belief' is meant to cover (p.29), but this is a mistake, as supposition does not have a mind-to-world direction of fit.[6] One does not go wrong by supposing that *p* when *p* is false.

Thirdly, to make 'belief' and 'desire' cover such a broad array of mental states forces us to include mental states that philosophers do not normally consider beliefs. *Perceptions*, for example, have mind-to-world direction of fit, so they would count as beliefs. However, most philosophers recognize that perceptions are importantly different from other mind-to-world states, and so they are cordoned off and not considered beliefs.

These sorts of considerations indicate that there is something wrong with the assumption of sparsity. It is standardly accepted that amusement is not a type of belief, imagination is not a type of belief, and perception is not a type of belief; although 'belief' is sometimes used as an umbrella term, it is usually intended to have a much narrower extension.[7] No one should be motivated to oppose non-doxasticism by

[6] In the way that Searle understands "direction of fit," at least.

[7] One *could* cling fast to 'belief' as an umbrella term, and hold that these really are all types of belief. Some research programs do treat perception as a kind of fragmented belief (Egan 2008, Gilbert 1991). However, this is not at all the standard way of speaking of belief in philosophy; if any mode of discourse

thinking of belief as such a broad psychological category that it subsumes all other informational states; this would also motivate one to be a doxasticist about perception.

What, then, distinguishes belief from perception, amusement, and imagination? How do we characterize a narrower conception of belief? I next turn to the attempts that philosophers have made to characterize this conception, and show that they are insufficiently precise to form the foundation of a fully-established theory that would trouble the non-doxasticist.

5.3.2 The Functional Role of Belief

If belief states are functional states, what is the functional role that makes for a belief state rather than, say, an imagining? This is, perhaps surprisingly, a seldom-discussed topic despite the compendious literature on the nature of belief. As is noted by Schwitzgebel, "[p]hilosophers frequently endorse functionalism about belief without even briefly sketching out the various particular functional relationships that are supposed to be involved" (2006).

Beliefs are usually taken to be an extremely general and basic sort of state. We have implicit and explicit beliefs, dispositional and occurrent beliefs, beliefs that are highly important to our social identities, beliefs that we obsess over, and beliefs in propositions that we might never consciously entertain because they are trivial or unlikely to ever bear on our lives, such as a belief that elephants do not have three beaks. Is there anything we can say to explicate such a general, wide-ranging sort of mental state? Attempts tend to be rough and illustrative, and many use terms that are in want of just as much explanation as 'belief'. For example, take Schwitzgebel's characterization in the *Stanford Encyclopedia of Philosophy* article on belief:

> Contemporary analytic philosophers of mind generally use the term 'belief' to refer to the attitude we have, roughly, whenever we take something to be the case or regard it as true (Schwitzgebel 2006).

---

is revisionary, this is.

The language of the characterization is merely suggestive, which Schwitzgebel knows; hence his use of 'roughly'. For example, what does it mean to 'take something to be the case' or 'regard it as true'? These phrases can't be taken in their everyday sense, for when someone assumes a proposition for the sake of argument, they take it to be the case or regard it as true. But assumptions are thought to be distinct from beliefs. You can assume that p by considering p without necessarily believing p. Moreover, there is a sense in which we take p to be the case when we perceive that p. Schwitzgebel's definition skirts very close to treating belief as an umbrella term, and it thereby has the previously-described fault of including all sorts of mental states that we do not typically consider beliefs. If 'taking something to be the case' is meant to be read in some other way, it's not clear that the explanans will be much better understood or less technical than the explanandum.

Despite the difficulty in characterizing the non-umbrella philosophical conception of belief, various analyses have been offered; it will be profitable to examine some of these. Cherniak (1986, p.6) considers a rudimentary theory that he calls an *assent theory of belief*: A believes all and only those statements that A would affirm. There are a number of problems with this sort of theory. "Affirm" here must mean "sincerely affirm" in order to allow for a person to make statements that she doesn't believe, such as when she lies or when she is acting in a play, but sincerity is not obviously a purely behavioral disposition—sincerity is worn in one's heart, not on one's sleeve—and it is difficult to further explicate the notion without resorting back to the notion of belief. Kaplan offers a more nuanced assent theory (which he calls an "assertion view of belief"):

> You count as believing P just if, were your sole aim to assert the truth (as it pertains to P), and your only options were to assert that P, assert that $\neg P$ or make neither assertion, you would prefer to assert that P (Kaplan 1998, p.68).

The most damaging problem with assent theories, for many, is that verbal affirmation is usually taken to just be one of many possible manifestations of a belief. Beliefs are responsible for the production of intentional action writ large, not just speech. A well-worn example: if A wants a soda, believes that going to the fridge can

result in the acquisition of soda, and no other overriding desires intervene, then *A* will go to get soda from the fridge. In addition, beliefs are thought to be responsible in the production of other mental states, not just speech and behavior. Given philosophical recognition of the behavior-guiding and reasoning-guiding role of belief, it's now common to think that verbal affirmation is not even a necessary feature of belief. A person with poor introspective ability might not assent to propositions that she believes because she does not realize what mental states are pushing her about in the world. Moreover, most philosophers would not want to join Davidson (1982b) in claiming that linguistic ability is necessary for an animal to have beliefs.

Russell subscribed to an internalized version of the public assent theory.[8] On this theory, *A* believes all and only those statements to which *A* would internally assent. To internally assent to a proposition is to entertain a proposition and have it be accompanied with a "feeling of assent." Cohen (1992) also holds a theory that is based upon phenomenal experience: having a belief requires being disposed to have a certain sort of "credal feeling." To believe that *p* is to be disposed to feel it true that *p*. We can call these *phenomenological theories of belief*.[9] The phenomenological theory is not popular. Credal feelings are fairly mysterious. My belief that I am currently in New York does not generate any strong or easily-identifiable credal feeling that I can discern. In any event, even if we do countenance credal feelings, the theory goes wrong for the same reason that the assent theory of belief went wrong; it doesn't consider the constitutive role that belief plays in intentional action. A person who felt it was true that *p* but who was never moved to act as if *p*, who verbally denied *p*, who never used *p* in reasoning, and so on, would not typically be said to believe that *p*.

More sophisticated accounts of belief attempt to tie belief more intimately to action. Braithwaite (1932) offers a simple version of such a theory, citing Bain (1859) as its original formulator: *A* believes *p* if and only if he or she is disposed to act as if *p* is true. Call this the *act-as-if theory of belief*.

What does it mean to act as if *p* is true? One natural interpretation is that it means

[8] This is pointed out by Cherniak (1986, p.6).

[9] Goldman (1993a) also argues for a phenomenological account of the attitudes.

to act as one would if *p* were true. However, a little reflection reveals that this won't do. Suppose that it is sunny outside and Sally wants to go for a run. She goes through the necessary motions by looking for her shoes and her headband, then doing some stretches. Now, how would Sally act if it were raining? This would depend on whether she knew that it was raining. Perhaps she's cooped up in her windowless office, so even if it were raining, she would still assume it was still sunny. In this case, if it were raining, Sally would nonetheless get ready to go running. But if Sally were to act as if it's raining (or if she believed that it's raining), she *wouldn't* get ready to go running. Therefore, to act as if *p* were true is not to act as one would act if *p* were true. It's difficult to avoid the conclusion that to act as if *p* is to act as one would if one *believed* that p were true. The act-as-if theory of belief isn't of much help, then; it is either a circular analysis, or it leaves us with an equally mysterious "act-as-if" notion.

The problems with analysis are not assuaged by turning to degrees of belief instead of flat-out belief. The literature on degrees of belief is rife with overly simplistic analyses of subjective probability. The most common is surely the *betting theory of degrees of belief*, advocated by de Finetti. The idea is that one's willingness to make monetary bets is treated as primitive.[10] Put aside the problem that such a theory assumes perfect rationality, and that a person will therefore always and only make bets that are in his or her favor. There are other problems with operationalizing the notion simply to dispositions to gamble. Many hold self-imposed prohibitions on gambling for social, personal, or religious reasons. Islam prohibits gambling, and so observant Muslims are not disposed to make any bets whatsoever. This should not be taken to imply that they have no beliefs. Eriksson and Hájek write that the literature on subjective probability contains "frequent gestures at some kind of betting interpretation, often accompanied with a slightly coy acknowledgment that this interpretation, well, isn't strictly speaking correct. But one is then immediately assuaged: at least it's

[10] "The probability P(E) that You attribute to an event E is therefore the certain gain p which You judge equivalent to a unit gain conditional on the occurrence of E: in order to express it in a dimensionally correct way, it is preferable to take pS equivalent to S conditional on E, where S is any amount whatsoever, one Lira or one million, $ 20 or £75" (de Finetti 1970, p.75).

'approximately correct', or 'correct over a significant range of cases', or a 'useful idealization', or what have you" (2007, p.184).

Why have so many imperfect analyses been offered? The key is in realizing that hedging is necessary. The problem with trying to identify belief with certain dispositions is that beliefs are massively *multi-track* dispositions: they manifest in a wide number of different ways in a wide variety of stimulus conditions. My belief that there is soda in the fridge might cause me to go get some soda; but then again, it might not; much depends on all sorts of other details about my cognitive state and the situation in which I find myself. I would also need to desire a soda, think that I'm not being rude to whatever guests I might have by getting a soda, be willing to suffer the effects of caffeine, be willing to ignore whatever diet I might be on, and so on. There are also all sorts of other behaviors or changes in cognitive state that might manifest the belief; it might be manifested when I make shopping lists, when I am asked whether there is soda in the fridge, when making plans for a party, and so on. Generalizations about the belief that there is soda in the fridge will, simply because of the sheer number of stimulus conditions and manifestation conditions, have to paper over the precise contours of the disposition by introducing hedging clauses, such as 'tends to', 'typically', or '*ceteris paribus*'. Goldman writes,

> One of the favorite sorts of platitudes offered by philosophers is something like, 'If x believes "p only if q" and x desires p, then x desires q'. But the relationship formulated by this 'platitude' simply does not systematically obtain. (Goldman 1989, p.79)

Therefore, philosophers who do attempt to describe the role of belief offer a few non-exhaustive and piecemeal generalizations with hedging clauses.[11] When the hedging clauses are left out, the characterization misses the extension that is clearly being aimed for; overelaborate characterizations that pile on necessary conditions are usually overly restrictive.[12] The reason for the paucity of descriptions of the role

[11] E.g. "Believing that performing action A would lead to event or state of affairs E, conjoined with a desire for E and no overriding contrary desire, will typically cause an intention to do A" (Schwitzgebel 2006).

[12] For example, see the characterization of belief offered by Graham and Stephens on page 56.

of belief likely has to do with the difficulty—verging on futility—of trying to characterize a massively multi-track dispositional state.

Belief has a certain behavior-guiding role, but we cannot *say* what its behavior-guiding role *is*, except in the loosest of terms. This is not just a problem for the non-doxasticist; it is a problem for anyone interested in developing a mature taxonomy of mental states. It should worry us that we cannot describe the role of belief except in very general terms and with impressionistic hand waves. A few non-exhaustive and imprecise generalizations, a coy admission that the account is only "approximately correct", and an elbow nudge that says "You know what I'm talking about, right?": there is a hope, in these maneuvers, that the reader already has a general idea of what the author means by 'belief'; any characterization offered is a nudge meant to help the reader triangulate on a concept that he or she already understands. We are deferred to a pre-existing folk conception.

This is especially problematic because the philosophical use of 'belief' invariably deviates from folk use. Note that the philosophical use of 'desire' deviates more obviously and more sharply from folk use than 'belief' does; 'desire' tends to be used to refer to sexual longing. This function is usually simply ignored: Dancy (2002, p.11) writes that 'desire' is used in philosophical discussion as "a term of art." If 'belief' in philosophical texts is a term of art that deviates from folk notions, we are owed a definition. But the functional role of belief is never adequately explained. Attempts to give an account of the role of belief always need to handwave back to the folk psychological notion. The criticism is akin to Austin's:

> [I]t is quite plain that the philosophers' use of 'directly perceive', whatever it may be, is not the ordinary, or any familiar use; for in that use it is not only false but simply absurd to say that such objects as pens or cigarettes are never perceived directly. But we are given no explanation or definition of this new use—on the contrary, it is glibly trotted out as if we were all quite familiar with it already.(Austin 1962, p.19)

Whatever one thinks of Austin's claim that 'directly perceive' has gone undefined, the form of the argument is valid. If the notion of belief in folk psychology deviates

from philosophical use, but philosophy turns to folk psychology to explicate the notion of belief, we have a problem. In the next section, I turn to the folk psychological notion, and argue that it is sufficiently inconsistent, variegated, and lush to support a number of different regimentations, including those that would reduce the role of belief in cognition quite dramatically.

5.4 'Belief' in the Mouths of the Folk

5.4.1 Folk Psychology

What is the folk psychological concept of belief? In "How To Define Theoretical Terms", (1970) Lewis offered a method for answering this question.[13] We collect all the "important" platitudes about belief that speakers are willing to assent to—i.e. propositions about belief that are "common knowledge"—conjoin them into a single theoretical statement, then *Ramsify*: that is, replace all instances of 'belief' with a variable, and identify the denotation of 'belief' with whatever unique entity in the

[13]Lewis's method is not the only method on offer. In fact, there are multiple sorts of answer that one might offer to the question "What is the folk psychological concept of *X*?", and Lewis's method gets at only one of them. It is worth doing a little ground-clearing here to stave off potential confusion. Because this is inessential, I do it in a footnote.

Firstly: the question "what is the folk psychological concept of belief?" might be asking about the content or meaning of the concept typically expressed by the word 'belief'. On the other hand, it might be asking about the role of BELIEF in a psychological theory that ordinary folk allegedly hold. (See Goldman (1993a) for further description of the difference between the two questions.) If you think that that the meaning of a term is not fixed by its role in a theory (say, you think that meaning is fixed by a causal chain that goes back to an initial dubbing), then the two questions will come apart. I am not currently interested in determining the meaning of the term divorced from its role in a theory. Most philosophers are meaning externalists of some ilk or another, and meaning externalism implies that the folk do not have introspective access to the contents of their concepts. We would need to do science to discover what those contents are. Therefore, because we are interested in how our folk understanding of belief might help us understand how to type mental states, we are interested in answering the second question, about the role that BELIEF plays in a folk theory.

Secondly: if we are interested in investigating the role that belief plays in a folk psychological theory, what sort of folk theory should we investigate? There are at least two different sorts of things that one can mean by 'folk psychology'. On one hand, the term can refer to an internalized theory that subserves our ability to attribute mental states to others. On the other hand, it can refer to a systematized collection of commonsense platitudes about mental states. Stich and Ravenscroft (1994) call the former an *internal* account of folk psychology, and the latter an *external* account. We are interested in the role of belief in external accounts of folk psychology. Again, we would need to do science to determine the actual tacit theory that people employ when attributing mental states (if that is how they attribute mental states), and these states could look very different from the ones that are suggested by folk words such as 'belief'. If we are going to turn to folk theory to help us type belief, we need to turn to our explicit statements about belief, and the principles that we can extract from intuitions about belief.

world satisfies the open sentence. If nothing in the world satisfies the sentence (perhaps because the platitudes are inconsistent), whatever *nearly* satisfies the theory, by being true of *most* of the important platitudes, is the referent of the term.[14]

This latter step is what causes problems. We very often use terms inconsistently and variously, and they need to be cleaned up in order to be made consisitent—certain platitudes must be discarded. To use Quine's term, out concepts and word use must be *regimented*. Simply by looking at platitudes about beliefs and the way in which the term is used, it is *not at all clear* what the commonsense concept of belief is, because there are multiple ways to regiment folk theory. The word 'belief' has a variegated texture and is used to perform a variety of functions, but some of these seemingly central functions are regimented away by philosophers or dismissed without argument as mere pragmatic effects.

What I want to argue is that there are enough tensions in the way that the folk use the word 'belief' to support an alternate regimentation: one in which 'belief' is ambiguously used to pick out two very different sorts of mental state. In other words, I want to show that something like the belief/acceptance distinction is nascent in our folk understanding of belief. The theory that I will propose in the next chapter has the consequence that many of our utterances are not expressions of belief at all. This *is* revisionary; but I have received complaints about it being *too* far from our folk notion. I want to show that it is not *that* revisionary. Sperber (2000, p.243) claims that the folk term 'belief' turns out to pick out two distinct mental states, like the common term 'jade' was shown to correspond to two different substances, jadeite and nephrite. In that case, however, we really had no idea that samples of jade could be either of two substances. There was nothing like a distinction in our external folk gemology. In the case of 'belief', traces of a bifurcated use can be discerned.

While still motivated by eliminative materialism, Stich made a case for the potential falsity of folk psychology by extracting the following principle from it:

---

[14] It might turn out that there are multiple entities that nearly satisfy the theory, by way of each satisfying a different subset of the platitudes. In such a case, there would be indeterminacy of reference. See Weatherson (2009) for a fuller account of Lewisian Ramsification.

> It is a fundamental tenet of folk psychology that the very same state which underlies the sincere assertion of 'p' also may lead to a variety of nonverbal behaviors. There is, however, nothing necessary or a priori about this claim that the states underlying assertions also underlie nonverbal behavior. There are other ways to organize a cognitive system. There might, for example, be a cognitive system which, so to speak, keeps two sets of books, or two subsystems of vaguely belief-like states. One of these subsystems would interact with those parts of the system responsible for verbal reporting, while the other interacted with those parts of the system responsible for nonverbal behavior. Of course it might be the case that the two belief-like subsystems frequently agreed with each other. But it might also be the case that from time to time they did not agree on some point. When this situation arose, there would be a disparity between what the subject said and what he did. (Stich 1983, p.231)

The separation between states responsible for verbal behavior and states responsible for nonverbal is a distinction that I will exploit in the next chapter. However, Stich here claims that it is a "fundamental tenet of folk psychology" that there is no such split. Is this true? Frankish calls the assumption that the mind does not keep "two sets of books" the *unity of belief assumption*, and he argues that the folk are not committed to it in his (2004, ch.2). It depends on how intuitions about belief, and the principles that underlie folk use of the word 'belief', are best systematized. A very strongly-supported systematization, I contend, would deny the unity of belief assumption: there are a number of striking divisions in the way that the folk speak of belief.

5.4.2 'Belief' in the Vernacular

How does folk usage of 'belief' differ from philosophical usage? Ordinary *talk* about the mental is lush, and teems with all sorts of different words for describing a person's inner life. Some philosophers have suggested that this shows that folk psychology is lush—i.e. it posits many different sorts of mental states—and that the various uses to which the word is put indicates that the folk concept of belief is not the same as the philosophers' concept.[15] This, however, is too quick. A rich lexicon of mentalistic terms does not necessarily imply a rich stock of mentalistic concepts. Two words can

[15] E.g. (Ratcliffe 2007).

share a meaning but have different pragmatic effects on the audience, such as is the case with 'sweat' and 'perspiration'. At this point, I want to describe patterns of folk usage without taking on theoretic commitments. I do not want to make any claims about whether any of the linguistic features I identify are due to the semantics of the word or the pragmatic use to which the word is put.

Needham (1972) is one of the few writers who has investigated use of the term 'belief'. He argues that 'belief' is a "peg word" that acts as a placeholder for a heterogenous range of various mental states. His main concern is admittedly to discover what anthropologists mean by the word, but he notes that the heterogeneity is even more varied and prevalent in ordinary discourse than it is in formal anthropological texts. I take it that a "peg word" is similar to Block's "mongrel concept"; to be a peg word is to be a word that refers indeterminately.

Take a word like 'democracy', 'love', 'settle', or 'terrorism'. It's at least prima facie unlikely that these words are used in a consistent way across an English-speaking population, or even by a single individual at different points in her life or in different contexts. These different instances of use will be intended to have different denotations. 'Love' will sometimes be intended by the speaker to refer to only romantic love, sometimes to filial love, and sometimes only strong liking, as when a boy says that he loves ice cream. The words exhibit some measure of indeterminacy or polysemy if not outright ambiguity, and some amount regimentation or explicit definition would need to take place before writing a technical article on any of these. 'Belief', I read Needham as claiming, is in the same boat.

Frankish (2004, 2009), has argued that the uses to which the word 'belief' is put reveals a number of inconsistencies and tensions. Beliefs are sometimes spoken of as conscious and sometimes as unconscious. The folk are acquainted with unconscious beliefs and drives being responsible for behaviors that we do not identify with. They are sometimes volitionally formed and other times forced upon us by circumstances. They are sometimes manifested primarily verbally, and at other times they are manifested through the actions that we take, and not what we say. We will sometimes take a person's avowed statements as an expression of their belief even if they

act contrary to their words; but on the other hand, it is not unnatural to tell someone, "I don't think you really believe what you're saying," and this isn't necessarily a charge of insincerity or lying.

The folk do use the term word 'believe' to pick out unconscious and non-volitional states with mundane content, although these are probably more often denoted with the word 'think' (as in, "Bob didn't take his umbrella…I guess he thinks it won't rain.").[16] However, the word is often used to serve other very core or prominent functions. Many of these functions are either ignored or regimented away by philosophers interested in belief. Consider the subtle differences between the following:

1. Obama believes that he's going to win the election.
2. Obama thinks that he's going to win the election.
3. Obama maintains that he's going to win the election.
4. Obama holds that he's going to win the election.
5. Obama is of the opinion that he's going to win the election.
6. Obama takes it to be the case that he's going to win the election.
7. Obama predicts that he's going to win the election.
8. Obama accepts that he's going to win the election.

All of these statements are of a family, but there are shades of nuance among them. A philosopher might treat these mostly synonymously, but to a writer or journalist, each would be appropriate in different contexts.

'Belief' is most prominently and paradigmatically used to refer to religious or spiritual belief. Searching for the word across various corpora, including *Google*, *Google Books*, and *The New York Times*, will reliably bring topics relating to religion

[16] 'Think' is also problematically varied in use, as it often is used to refer to acts of occurrent, conscious thought. This is presumably why philosophers settled on the label 'belief' instead of 'think' in discussions of belief-desire psychology. It doesn't always refer to conscious acts of thinking, however; it also has a reading that refers to a dispositional standing state.

to the forefront. If you mention 'belief' to a non-philosopher, you will bring to their mind the notion of "belief in God" or "belief in a higher power." Note that these locutions do not employ propositional attitude verbs. "Believing-in" sentences are syntactically different from those that include the propositional attitude of "believing-that", and it has two functions: a commendatory function, in which one attests confidence in the object (as in, "You can do it! I believe in you!") and a function of expressing existential commitment (as in, "Herbert believes in ghosts.") (Macintosh 1995).[17] (Interestingly, 'confidence' can also be used to generate "confidence-in" and "confidence-that" sentences, but "confidence-in" does not have the function of expressing existential commitment.)

Statements of belief in God may serve both of these functions simultaneously. To assert one's faith by saying 'I believe in God' is not necessarily to simply make the claim that God exists; it can also be an offering of praise. Because these functions can be distinguished but can also both be manifested in a single speech act, they can easily be confused with one another. A very adamant statement of faith might be used to strongly express praise for God; this shouldn't be taken to imply that the speaker has absolute conviction that God exists. If the functions are separable, a strong commendation does not necessarily imply a similarly high degree of confidence in the existence of the thing being commended.

While these functions may seem associated only accidentally, there is evidence that their co-occurence was seminal to the development of the English word. The etymological ancestor of 'believe' is the Proto-Germanic '*galubjan*', which means *to esteem, to hold dear*, and which contains the root '*leubh*', meaning *to love*. Given that we are now used to drawing a bright line between the cognitive and the conative, it is surprising to find the conative so central to the semantic ancestry of 'belief'. The commendatory function of the word 'belief' is a crucial feature of the word's ancestry, not an parasitic use that got tacked on at some point.

---

[17] A joke that plays on the ambiguity of function: Santa Claus, lying on a couch, says to his psychiatrist, "my parents never believed in me."

"Belief-that", on the other hand, does not obviously have the same conative function that 'belief-in' has. Nonetheless, typical use of 'belief' as a propositional attitude verb does depart from typical philosophical use, and I do think a plausible case can be made for 'belief-that' often having something akin to the commendatory function of "belief-in."

'Belief' often conveys uncertainty. Ratcliffe (2007, p.188) points out the difference between "careful, there's a lion in the room," and "careful, I believe there's a lion in the room." The latter voices a suspicion rather than a statement of fact. This sort of hedging that comes from using the word 'belief' can be an indication of low confidence, but not always. More precisely, the word 'belief' indicates something about the status and quality of the evidence that supports the belief. It conveys that the mental state is unsupported by epistemic reasons, or at least partly held for reasons other than epistemic reasons. For instance: the song *I Believe I Can Fly* does not translate well into *I Think I Can Fly*. This isn't because the former title demonstrates less confidence than the latter—if anything, the opposite is true. It is because the latter does not communicate the aspirational motivations that support the conviction.

Consider the slight difference between "Obama thinks he's going to win the election" and "Obama believes that he's going to win the election." When would a journalist would use one sentence rather than the other? The former has Obama making a prediction; the latter connotes that something like hope is supporting and maintaining his prediction. This isn't to say that he doesn't also have good evidence—he might know what the polls say and might be entirely justified in predicting that he'll win. The non-epistemic factors might not even raise Obama's degree of confidence to a level that cannot be justified by evidence. Rather, it is to say that the holding of the belief is causally overdetermined: it is maintained by both evidence and pragmatic factors. Just as a supporter at a rally might shout 'I believe in you, Obama!' to express a commendation, she might shout 'I believe that you'll win, Obama!' to express that her prediction is at least partly based on approval. There are, perhaps, traces of the conative that still reside in 'belief-that'.[18]

[18]This isn't to say that 'belief-that' necessarily, or even typically, connotes hope or commendation.

These various eccentricities of the word 'belief' are partly responsible for why a nervous swimmer might ask "Do you think there are sharks in these waters?", but would not ask "Do you believe there are sharks in these waters?" The word 'belief' is typically reserved for cases in which the belief is held for reasons other than being supported by evidence, or for cases in which the evidence cannot be brought to mind.[19]

Thus, there seems to be two reasonably natural readings of 'believes', one which is close in meaning to 'thinks', and one which is close in meaning to 'has faith', and the core use of 'belief' seems to be the latter. Mixing the two can produce zeugmatic effects: consider "Joe believes that his car is parked on the west side of the street and that Jesus saves." It sounds strange. This may just be due to the inanity of expressing such a thought, but there is enough material here, I believe and think, to support the contention that there are two distinctly different functions of the word 'belief', and a regimentation of folk psychology should respect that division—particularly because, as I'll argue, a mature psychological theory should include a similar division.

5.4.3 Semantics and Pragmatics

Should these observations trouble those who think that the folk are committed to the unity of belief assumption? One might remain unperturbed, and hold that the faith-like features of 'belief' identified above are attributable to pragmatic effects rather than anything having to do with the lexical semantics of the word. Only certain inferences that we are willing to draw from statements about belief are due to semantical

---

Rather, it connotes that the mental state is partly held or maintained for non-epistemic reasons. Hope and commendation are just two likely ones. "Obama believes that he will lose the election" doesn't express that Obama hopes he'll lose the election, but it does suggest that his prediction is formed and supported in part by something other than just evidence from the polls... something like a pessimistic or defeatist character, perhaps.

[19] It's an open question whether words in other languages that are typically translated into 'belief' have the same function. The distinction between 'croire' and 'penser' in French, at least, seems to be roughly the same as the distinction between 'believe' and 'think' in English. Different dialects of English might also treat the word differently; British English seems more tolerant of using 'believe' to mean 'think'. I am happily willing to admit there can be important cross-cultural differences in these matters.

relations between belief and other concepts, and only certain infelicities are semantical infelicities. There are two possible explanations for the discrepancy between philosophical use and the everyday use. Perhaps in many instances the word 'belief' picks out a slightly different concept in ordinary English than it does in philosophical discussions. Philosophers have reappropropriated the word for their own purposes; a reader unfamiliar with the appropriation might have to struggle to figure out what the philosopher means by his unnatural use of the word. On the other hand, perhaps the concept picked out by the word affords the same semantical inferences in both philosophical conversation and everyday conversation, but the subtle nuances ignored by the philosopher are conversational implicatures that are not part of the word's meaning. The richness of the English lexicon allows for many roughly synonymous ways to express our mental states, so 'belief' is chosen in certain contexts because of the speech act effects that it generates. The former explanation appeals to a difference in semantics, and the latter appeals to a difference in pragmatics.

It's important not to ignore this distinction. Some have claimed that because philosophers fail to consider the various subtleties of ordinary mental talk, philosophical belief-desire psychology diverges sharply from folk psychology. These criticisms do not always consider that the subtleties might just be a part of word's pragmatics.[20] It would be too hasty to conclude that because philosophers use the term differently than it appears to be used in the vernacular, they mean something else by the term. It would also be too hasty to make an argument for non-doxasticism that relies entirely on folk usage of the word.

Manfred Spitzer makes this error when he attempts to make a very quick argument for non-doxasticism. He writes, "patients rarely say that they "*believe* that (such-and-such)," but rather state that they "*know* that (such-and-such)." That is to say, delusional patients express conviction and certainty about certain statements rather than admit that these statements may be subject to discussion and inquiry"

[20] E.g. see Ratcliffe (2007, ch.7)

(1990, p.381). This argument assumes that belief requires a certain amount of uncertainty. However, if a patient says, "I don't *believe* that I'm being followed; I *know* it," it is very plausible that they are pragmatically conveying that their belief is not a *weak* belief or *mere* belief. Consider the following exchange:

A: "You spent half a year's salary on a car?"
B: "A *car*? This isn't a *car*! This is a Lamborghini!"

Despite his literal words, Speaker B doesn't think that a Lamborghini is not a car; his paralinguistic emphasis conveys that it isn't a *mere* car or a *typical* car.[21]

So, perhaps the various nuances of the word 'belief' are attributable to pragmatic intent; perhaps the semantic inferences that people are willing to draw from 'belief' are precisely those that the philosopher would wish to draw. 'Belief' only appears to refer to a heterogeneity of mental states because pragmatic effects draw the listener's attention to only some subtypes of belief (such as "mere belief"), and the use of 'belief' to pick out a state formed by something other than evidence might just be a type of pragmatic implicature.

However, none of this can simply be assumed. Tests exist to determine whether various linguistic effects are semantic or pragmatic, but as far as I know, they haven't been done. There are plenty of books and papers that purport to be about the semantics of belief, but these typically investigate the semantics of any attitude verb that generates an intensional context; they don't investigate the semantic differences, if any, between various sorts of intentional attitude verbs, such as 'believe', 'desire', 'think', and 'hope'. It is optimistic to think that all the various uses to which the word 'love' or 'democracy' are put will reveal a stable and determinate lexical semantics. I doubt that anyone thinks that we can develop a single consistent theory of 'love' or 'democracy' that *perfectly* captures how the word is used in the vernacular, even if we take speech act effects into account; people simply have too many inconsistent ideas

[21] Other examples of this phenomenon in the wild include: *(1)* HBO's slogan, "It's not TV. It's HBO." *(2)* A scene in the movie *Pulp Fiction* in which Bruce Willis's character's girlfriend asks him where he got his motorcycle. He responds, "It's not a *motorcycle*, baby. It's a *chopper*." *(3)* A scene in the movie *Love Actually* in which a store clerk says, "This isn't a bag, sir. This is *so* much more than a bag."

about love and democracy. If Needham is right that 'belief' is an indeterminate peg word, we won't find stability here either. The onus is on the philosopher who thinks there is a consistent lexical semantics to tell us what the lexical semantics looks like.

5.5 A Subtype of Belief, or Non-Belief?

In the following chapter, I will present an attitude—acceptance—that roughly corresponds the to the attitude that expresses faith and commendation and that is formed for non-epistemic reasons. Ordinary language and folk vernacular use reveals that the folk have a nascent but incomplete grasp of the division between belief and acceptance.

One might think that I am doing harm to my claim that acceptance (and hence, delusion) is not belief. I've just argued that one of the core functions of the word 'belief' is to pick out acceptance. So, why think that acceptance is not just one type of belief? Why not think that, just as we discovered that there are multiple sorts of memory, we discover that there are multiple sorts of belief?

To some extent, this is just a terminological dispute. If someone were adamant that acceptances deserves the name 'belief', given that ordinary English calls them this, I would only be willing to put up so much of a fight. What is important is describing the roles of the various psychological types that we should posit; deciding on labels to affix to those roles is of secondary importance. However, I do think that once the revisionary part of the project kicks in, and we start looking at the notion of acceptance, there are two considerations that militate strongly in favor of refraining from calling acceptances 'beliefs'.

Firstly, doing so shrouds important differences between acceptance and the belief state that philosophers typically talk about. Although the folk indiscriminately use the word 'belief' to refer to either beliefs or acceptances, close investigation of acceptance reveals that it is not as similar to belief as one might think. There might, in fact, be no natural supertype that includes both types of state; it would be grueish

and gerrymandered. Calling the various sorts of memory 'memory' was not as problematic. Given that a core function of 'belief' in the vernacular is to refer to acceptance, it might have once been appropriate to bestow the title on acceptance, and use a word like 'thought' to refer to the other state, but since philosophers have already adopted the word to name that other type of mental state, they can have it. Secondly, and more importantly, a close investigation of acceptance reveals that certain subtypes of acceptance are states that we do not traditionally consider beliefs. Namely: *assumptions* and *suppositions* are acceptances. If acceptances are beliefs, then so too are assumptions. This is probably too big of a bullet to bite.

5.6 Summary

'Belief' in the mouths of philosophers is usually presented as a term of art with a technical use, but because of the difficulties in actually defining this term of art, they often handwave back to a folk notion. However, the folk notion of belief (at least if read off of the way the word 'belief' is used) serves the function of denoting two importantly different types of mental state. Only one of these is that which philosophers typically have in mind.

In the next chapter I present the attitude of acceptance, and pull it into a theory that can explain psychopathological delusion.

Chapter 6
A Positive Proposal

A distinction between belief and acceptance has been drawn by many philosophers, including Bas van Fraassen (1980), Robert Stalnaker (1984), Michael Bratman (1987), George Rey (1988), Jonathan Cohen (1992), and Pascal Engel (1998, 2000), among others. I also include in this category Ronald de Sousa's "assent" (1971), Daniel Dennett's "opinion" (1978b), Dan Sperber's "reflective belief" (2000, 1996), and Keith Frankish's "superbelief" (2004). These writers do not draw the distinction between belief and acceptance (or whatever their preferred terminology) in precisely the same way, but they tend to either cite the earlier writers as influences and precedents, or they comment on the similarities.[1] It appears that everyone is circling around a common idea.[2] This idea is one that can be fruitfully used to explain pathological delusions.

I will proceed as follows. It is easiest to get an initial grip on the notion of acceptance that I have in mind by considering various examples, so I'll begin with a battery of these, and provide some provisional and preliminary characterizations. In the sections after, I'll tighten the characterization of acceptance and describe its formation conditions, its effects on behavior, and its role in reasoning, by detailing three defining features:

ROLE IN REASONING: Acceptance that p is a disposition to use p as a premise in conscious, deliberative reasoning.

---

[1] See Engel (2000) or Frankish (2004, chap.3) for a survey of various accounts of acceptance, including those of the writers listed above.

[2] For ease of exposition, I'll speak of each of these authors as having a theory of acceptance, even though not all would endorse the term. When I speak of Dennett's theory of acceptances, for example, you can mentally replace the word with 'opinions'.

FORMATION CONDITIONS: Acceptances, unlike beliefs, can be, but are not necessarily, volitionally formed.

BEHAVIORAL EFFECTS: Acceptances, unlike beliefs, primarily affect verbal behavior.

I then show how treating delusion as a form of acceptance allows us to explain the features of delusion that are in want of explanation.

Finally, I turn to questions about the formation of delusions. If delusions are acceptances, then why do delusional individuals ever accept such bizarre contents in the first place? The account I propose explains delusional acceptances as the result of pathological *cognitive feelings*. In the latter half of this chapter, I explain what cognitive feelings are, and present evidence that they are inculpated in the formation of delusional acceptances.

6.1 A Preliminary Description of Acceptance

It is easiest to get a grip on the role that acceptance plays in reasoning by considering a type of acceptance with which everyone is already familiar. *Supposition* and *assumption* are subtypes of a broader category of acceptance; *assuming* or *supposing* that *p* are ways of accepting that *p*.[3] However, suppositions are easily and readily discharged: there comes a point where we divest ourselves of the supposition and are no longer disposed to use it in reasoning about a certain topic. This is not the case with all acceptances. We can think of acceptances as suppositions that can be more full-bodied and long-standing: ones that might be used as premises in reasoning at any point in time, in many different contexts, and might be enfolded into our personal narratives about ourselves. The two following examples illustrate the difference:

THE TRIAL LAWYER: A trial lawyer needs to defend a client against a murder charge. She has excellent evidence that her client is guilty; in fact, she thinks that he probably is guilty. However, she is paid to muster the best defense that she can.

[3] I use 'assumption' and 'supposition' interchangeably from here on out.

> For the sake of her profession, she *assumes* that her client is innocent and works from there. Her client's innocence is used as a premise in future reasoning in the context of the courtroom.

It's important to note that when we assume a proposition, we do not necessarily form a new belief, even temporarily. Assumptions are not merely low degrees of belief: you can assume something while not believing it in the least. When in the courtroom, the lawyer bears a qualitatively different sort of attitude toward the proposition that her client is innocent than she does toward the proposition that her client is guilty. Assumptions are also not merely temporary beliefs. There is nothing especially controversial about the existence of an assumption for the sake of argument or a conversation, and we all recognize that it's possible to assume something without believing it. It is also not necessary for a person to be disposed to publicly admit that their assumption is a mere assumption and not a full-fledged belief. As a lawyer in a courtroom, I would not admit to disbelieving the conclusion of my own arguments (at least, doing so would be doing a severe disservice to my client).

Now consider a similar example:

THE MURDERER'S SON: A man's father is accused of murder. The son has excellent evidence that his father is guilty. If he were to think coolly on the subject, he would recognize that he does in fact suspect that he probably did it. However, for the sake of some (potentially misguided) familial duty, he plants his foot down: his father is innocent. He forms a commitment to his father's innocence and treats this as a premise in future reasoning whenever the topic comes up, whether inside a courtroom or not.

We typically think that when a person is adamant about a certain subject, and will brook no evidence or argument to the contrary, their assertions and behaviors manifest their beliefs. However, these sorts of assertions and behaviors can, in certain contexts, be indicative of assumption, not belief. Just as the lawyer might vociferously attest his client's innocence to a jury, even getting angry or riled up in the face of opposition, the murderer's son might similarly vociferously get angry with friends

who try to prod him into admitting his father's guilt. Given that we already countenance supposition, why should we not think that some suppositions could swamp a person's entire life and never be discharged? Remember that suppositions are not merely short-lived beliefs: it is not their being temporary or context-bound that fundamentally distinguishes them from belief. Their distinguishing features reside elsewhere. Therefore, why not think that just as there are both temporary suppositions and temporary beliefs, there are long-standing suppositions that are distinct from long-standing beliefs?

A noticeable feature of the trial lawyer example is that the lawyer knows that he is merely assuming that his client is innocent. I said above that a lawyer need not be disposed to publicly admit that his assumption is a mere assumption. A lawyer is typically, however, disposed to *privately* admit that his assumption is a mere assumption. One might think that this is a crucial feature of supposition that distinguishes it from temporary belief. That is, one might think that supposition requires a kind of metacognitive introspective awareness of one's mental states: a person's supposition ceases to be a supposition if it is confused with a belief.

However, this does not seem to be true. You might suppose that p for the sake of argument, and in the midst of a heated debate, temporarily lose track of the fact that you have only supposed that p and that you don't actually believe it. In fact, I think it is likely that in the case of long-standing acceptances, it is common to be unaware whether you actually believe p or merely assume it. Consider first the two following cases:

THE CAREER-BUILDING PHILOSOPHER: A graduate student in philosophy finds a passage in which a prominent philosopher makes an argument for realism with an unargued premise p that is not beyond reproach. He realizes that denying this premise could form the cornerstone of a version of fictionalism that no one in the literature has propounded. He is sitting on a gold mine of potential articles! So, the student starts publishing a series of articles that show what follows from

a denial of p, eventually using not-p as premise of his own in many of his arguments. He gives talks on his version of fictionalism; he acts as if he has a high subjective probability in it, he is willing to bite the bullet in print and treat anything that conflicts with not-p as something to be rejected. Many years later, after having established a career, an old friend at a bar asks him if he really believes this stuff. The fictionalist says, in confidence, what he has known all along: "No. But it's a position out there in logical space worth investigating. And it got me tenure!"

THE UNBELIEVING CLERGYMAN: An observant Catholic youth has his faith shaken in his senior year of high school and begins to harbor doubts about God's existence. He had always planned to become a seminarian upon graduation, and he still plans to do so, partly in a hope that an environment of that sort will address his questions and restore his faith. He enters seminary, and as he goes through, he never explicitly voices any skepticism, always outwardly exhibiting an unswerving belief. God's existence is taken as a premise in all of his studies; nonetheless, he comes to believe that God does not exist. Eventually, he graduates, and following his best career prospects, becomes a pastor. Though he does not believe in God, he sees goodness in continuing to preach, and when members of his congregation ask him for advice in matters of faith, he reasons from a supposition that God exists. Although he feels guilty, he keeps up a charade.[4]

---

[4]Cases such as this are not uncommon. Dennett and LaScola have conducted a number of interviews with members of the clergy who have lost their faith or have recognized that they never truly believed. I drew from these interviews in devising the above example, such as an interview of one anonymous clergyman, who reports,

> I didn't believe in God, but I thought, before I reject the street version of Christianity, I'll go to seminary for a year. And I'll argue with the best theologians and the best religious scholars, and then I'll get out. I'm not going to leave the church; I'm not going to leave what I was formed in until I have a chance to confront the scholars and argue and see what's going on. Is there anything in this God business? That was the way I kind of put it. Is there anything to this? So I determined to enter seminary ... My first few years of doing this were wracked with, God, should I be doing this? Is this—? Am I being—? Am I posing? Am I being less than authentic; less than honest? ... And, I really wrestled with it and to some degree still. But not nearly as much. (Dennett and LaScola 2010, p.129)

In each of these cases, a person lacks a belief, but supposes or assumes that it is true for the purpose of constructing arguments, giving advice, and guiding their interactions with others. They are not dissimilar from the lawyer, but their suppositions are not confined to a single courtroom—they have folded that which they accept into their identities and lives. What they are doing is akin to engaging in a pretense of belief. One might think that they are criticizable for this; they are acting in bad faith. They know that that what they privately believe and outwardly avow do not line up.

However, imagine that they had poor introspective abilities. Imagine two biographies that are essentially the same, but in which the protagonists do not recognize that they are engaging in something like pretense (the sections that differ from the vignettes above are in italics):

THE SELF-DECEIVED PHILOSOPHER: A graduate student in philosophy finds a passage in which a prominent philosopher makes an argument for realism with an unargued premise *p* that is not beyond reproach. He realizes that denying this premise could form the cornerstone of a version of fictionalism that no one in the literature has propounded. He is sitting on a gold mine of potential articles! So, the student starts publishing a series of articles that show what follows from a denial of *p*, eventually using not-*p* as premise of his own in many of his arguments. He gives talks on his version of fictionalism; he acts as if he has a high subjective probability in it, he is willing to bite the bullet in print and treat anything that conflicts with not-p as something to be rejected. Many years later, after having established a career, an old friend at a bar asks him if he really believes this stuff. *The fictionalist says, (and he fully believes that what he says is true), "Of course I do!"*

THE SELF-DECEIVED CLERGYMAN: An observant Catholic youth has his faith shaken in his senior year of high school and begins to harbor doubts about God's existence. He had always planned to become a seminarian upon graduation, and

> he still plans to do so, partly in a hope that an environment of that sort will address his questions and restore his faith. He enters seminary, and as he goes through, he never explicitly voices any skepticism, always outwardly exhibiting an unswerving belief. God's existence is taken as a premise in all of his studies; nonetheless, he comes to believe that God does not exist. Eventually, he graduates, and following his best career prospects, becomes a pastor. Though he does not believe in God, he sees goodness in continuing to preach, and when members of his congregation ask him for advice in matters of faith, he reasons from a supposition that God exists. *The sense of guilt he feels when he reflects on his skepticism becomes overwhelming, and he resolves this tension by never again reflecting on his disbelief, always avoiding situations in which he would be forced to appraise God's existence. If he asks himself whether God exists, he will reflexively answer in accord with his acceptance—"of course!"—and quickly redivert his attention.*

In these cases, the mental state in question looks and overtly behaves an awful lot like belief; but all that I have changed between the two philosopher cases and the two clergyman cases is a metacognitive, introspective awareness of a state that is not belief. Introspective confusion about what you are assuming for the sake of argument is not enough to render your assumption a belief. If the pastor will not admit to himself that God does not exist, can we really conclude that there is any sense in which he really doesn't believe that God exists? Yes. Various triggering conditions could reveal, to his surprise, that he does not believe what he thinks he believes. For example, he could discover, on his deathbed, that it does not matter to him whether he receive last rites.

The attitude that is exhibited in all of these cases is not belief, but acceptance. The trial lawyer accepts that his client is innocent, but only when he is in the courtroom. The murderer's son accepts that his father is innocent in a much wider variety of contexts. The self-deceived philosopher and clergyman have acceptances that they do not recognize as something that they can discharge.

Suppositions are states that exist momentarily for the purpose of reasoning and are quickly discharged. However, by reflecting on the way that what begins as a simple assumption for the sake of argument can balloon outward and swallow one's life and one's identity, and then considering the way that we can then lose track of what we genuinely believe and what we merely suppose, we can see that fleeting assumptions are only a narrow subtype of a more general mental state that is responsible for a lot of our assertions and behaviors.

Acceptance, and particularly acceptance without belief, has been used to explain a wide variety of phenomena, including:

- Self-deception.
- Apparent beliefs that a person volitionally adopts for instrumental purposes.
- "Quasi-beliefs", in which a person claims to believe the content of a sentence that is meaningless or that they do not actually understand.
- "Flat-out beliefs," as opposed to graded beliefs that admit of degree.
- The attitude that scientists and philosophers hold toward their academic theories.
- Cases in which a person's behaviors reveal that they do not really believe what they say.
- Certain superstitions, mystical attitudes, and religious convictions.

The above characterization is meant to help the reader grasp an intuitive notion of acceptance. In what follows, I sharpen the notion by describing various features that importantly distinguish acceptance from belief. Doing so will give the attitude the shape of a robust theoretical posit.

6.2 Features of Acceptance

6.2.1 Acceptance and Reasoning

Different theories focus on different aspects of acceptance as the most important. On my conception of the state, the defining feature has to do with its role in reasoning, just as the most defining feature of assumption is its role in reasoning. When completing a logic problem, I might be told to assume that *p*. Doing so makes it available for use as a premise in the proof. When with friends at a dinner party, I might assume for the sake of facilitating conversation that they are right about *q* even though I don't believe *q*. This allows me to tease out the implications of *q* in discussion to come.

Jonathan Cohen has produced one of the more influential and seminal accounts of acceptance in his (1992). According to the definition that he offers, to accept that *p* is "to have or adopt a policy of deeming, postulating, or positing that *p* — i.e. of including that proposition or rule among one's premisses for deciding what to do or think in a particular context" (Cohen 1992, p.4). There are inadequacies with this definition—for instance, what does it mean to include a rule among one's premises?—but the sentiment behind it is clear enough. To accept that *p* is to have a certain disposition to use it in reasoning. Frankish refers to this as the "premising conception" of acceptance (2004, p.81).

Of course, the contents of belief are used in premises in reasoning as well; this description of the role of acceptance requires some refining in order to distinguish it from that of belief. Try for yourself to articulate the difference between supposing that *p* and believing that *p*—it's not easy. (The problems here are not dissimilar from the problems faced by the philosophers in the previous chapter tasked with describing the dispositional role of belief.) Nonetheless, it is possible to make progress. Acceptances are manifested in only a particular sort of reasoning: conscious, deliberative, Type 2 reasoning.

6.2.1.i Type 2 Reasoning and Simulated Belief

A very old debate in psychology concerns whether human reasoning should be modeled as we tend to experience it: as a deliberate and sequential manipulation of thoughts. An opposing picture casts human reasoning as the product of many associative processors running in parallel. It is now widely acknowledged by psychologists that there is some truth to both pictures. *Dual process theories* posit two distinct types of reasoning processes operative in human cognition, each of which is performed by different cognitive mechanisms (Sloman 1996). The first sort of processes are performed by evolutionarily older systems that are shared with many animals. These operations are typically thought to be fast, parallel, nonconscious, automatic, and associative. The second sort of processes are evolutionarily younger and are performed by systems that are uniquely human; these processes are slow, conscious, serial, rule-governed, and domain-general. The older, quicker, less rule-governed reasoning processes are known as Type 1 processes. The younger, slower, deliberative processes are Type 2 processes.[5] When we employ deliberate, conscious reasoning, we employ Type 2 processes. Variants of the distinction were independently proposed by researchers studying implicit learning (Reber 1993), deductive reasoning (Evans 2009), heuristics and biases (Kahneman et al. 1982), and social cognition (Smith and Collins 2009). Sometimes, our quick-and-dirty Type 1 processing systems will yield a first-glance verdict on a certain question; when we sit and contemplate on the issue, we come to another conclusion. Type 2 reasoning allows us to overcome the mistakes that we reflexively make when reasoning about probabilities, for example: deliberation lets us avoid committing the Gambler's Fallacy.

The function of the Type 2 reasoning system (which I will refer to, as Sperber (2000) does as an "inferential device") is to generate new beliefs that are logical consequences of our other beliefs (Recanati 2000). However, we are able to exploit it for

[5] Psychologists sometimes refer to this literature as the "dual systems" literature, and speak of the Type 1 system and Type 2 system in order to refer to two reasoning systems. I suspect there is a motley collection of subsystems responsible for Type 1 reasoning, each with their own individual functions, so I prefer to speak of dual process theories rather than dual systems theories.

other purposes. We are able to use Type 2 reasoning to draw inferences from propositions that we consider but don't actually believe. This makes acceptance a form of simulated belief. Simulation exists when a mechanism that has a certain function in a system is used offline—i.e. it is detached from that system and used for a separate function (Nichols et al. 1996, Recanati 2000).

Acceptances are therefore similar to what Goldman (1993b) and Nichols et al. (1996) call 'pretend beliefs'. To pretend-believe p is to feed a representation of p into the inference device without necessarily believing it in order to enable counterfactual reasoning. The general shape of this suggestion is correct, but calling them "pretend" beliefs invites criticism that can be avoided. Acceptance and pretense both use inference device offline for their own purposes, and so are both forms of simulated belief, but despite what is sometimes claimed,[6] acceptance is not pretense. Firstly, many regular beliefs that conflict with the entertained proposition are not screened off from inference with acceptance in the way that they are screened off from inference pretense; using an acceptance in inference allows us to (nearly) freely draw from our other beliefs. Secondly, acceptances are not elaborated or fancifully embellished in the way that pretenses are. Thirdly, pretending that p usually causes appropriate pretend behavior that is lacking when one assumes that p.[7]

If we are in a restaurant and I tell you to pretend that I'm a Wild West cowboy (and you take me up on my invitation), you will also pretend a whole host of propositions. You might pretend we are no longer in a restaurant, but a saloon. You might pretend to be a cowboy yourself and affect a drawl. On the other hand, if I ask you to *assume* that I am a cowboy, you won't engage in this sort of play (unless you misunderstand my request for an invitation to pretend). Instead, you might ask how a Wild West cowboy has come to be in twenty-first century Brooklyn. If I tell you to pretend that a banana is telephone, you'll pick it up and talk into it; if I ask you to assume that this banana is a telephone, you might do this, but then ask me how a banana could

[6] E.g. "When someone assumes that p [...] the agent unmistakeably *pretends to believe*" (Recanati 2000, p.287).

[7] These characteristics of pretense are drawn from (Nichols and Stich 2000).

possibly be a telephone. The belief that it is not a telephone might be screened from entering inference, but associated beliefs, such as the belief that it is not a banana or the belief that no banana is a telephone, must be individually screened off through other acts of acceptance. You also will not be motivated to behave in a way that accords with having a telephone in front of you; you won't pick it up.

On an influential model of pretense (Nichols and Stich 2000), pretense involves the construction of a "possible world" representation. A "possible world box" stores representations of the way the world could be if a set of initial premises were true. As the pretense goes on, it is embellished, and beliefs that conflict with the possible world representation being constructed are screened from being used in the inference device to further elaborate the possible world. Acceptance does not involve anything like a possible world box that is elaborated on and that screens beliefs. Accepting a proposition disposes one to use it as a premise in reasoning.

Recognizing that acceptance is a form of simulated belief in addition to pretense opens up new avenues for explaining mental phenomena that have been traditionally explained with pretense. For example, it is sometimes said that counterfactual reasoning employs simulated belief. A counterargument that has been made is that autistic children have problems indulging in pretend play, but they are able to employ counterfactual or suppositional reasoning, so counterfactual reasoning cannot involve simulation (Nichols and Stich 2000, Scott et al. 1999). This argument is only valid with the additional premise that pretense is the only form of simulated belief. Autistics may have difficulties screening off relevant beliefs, embellishing possible world representations, or engaging in pretend behavior, but still be able to accept propositions and use them in offline reasoning.[8] Gendler has proposed that self-deception involves pretense, but I think it is more likely that it involves acceptance, as self-deception does not always involve embellishment or fanciful playacting.

I will not follow up on these suggestions, as my focus is on delusion. I do, however, note that the difference between pretense and acceptance strongly supports

[8]It might be said that pretense involves acceptance and has these other additional features as well.

an acceptance theory of delusion over a pretense theory of delusion. If a woman says to her husband, "pretend for a moment that I'm not your wife", he might make-believe a new story about lives and engage in a flirtatious play-act with her, saying "Helloooo... what's your name?" If she says, "assume for a moment that I'm not your wife," a more natural response would be to say, "OK... but then who are you, and who have I been living with for the past few years?!"

6.2.1.ii Validation

The inference device produces representations that are sometimes admitted into our store of beliefs, but at other times are not. We can reason about our beliefs and thereby form new beliefs, but when we assume a proposition for the sake of argument, we are not forced to thereby believe the consequences that we draw from it. The consequences are insulated from our belief store. Sperber (2000) and Recanati (2000) propose that the outputs of the inferential device can enter our belief store by being *validated*. The basic idea is that the conclusion to a Type 2 inference will become the content of a new belief only if all of the premises are believed. Inputs to the inferential device are tagged as either validated or unvalidated depending on their provenance: those that come from our store of beliefs are validated, and those that are *merely* accepted or pretended or imagined are not. If all of the premises used by the inferential device are validated, then the output is validated. If even one of the premises is unvalidated, then the output is unvalidated.

Only validated propositions can be integrated as new beliefs. Whatever Type 1 mechanisms are responsible for belief revision treat validated outputs of the inferential device as evidence, and conditionalize upon them or otherwise use them to update the belief store.[9] Type 2 reasoning itself does not tell us anything about how to revise our beliefs. It's often been noted that deductive arguments tell you which propositions follow from other propositions, but not whether a conclusion should

[9]Conditionalization would require the outputs of the inferential device to be weighted. The weight of the output might be dependent on the weights of the inputs to the inferential device in some way, but the weights would not be themselves used in the inference.

be accepted or whether a premise should be rejected. (Thus the familiar phrase: one man's *modus ponens* is another man's *modus tollens*.) Bermúdez (2001) distinguishes between procedural rationality and epistemic rationality: procedural rationality consists in reasoning in ways that conform to formal logical laws of consistency, and epistemic rationality consists in updating one's beliefs in accord with norms of good reasoning. He then argues that while schizophrenics are perfectly procedurally rational, they are pathologically epistemically irrational in that they do not reject (as they should) the bizarre conclusions that they draw from their pathological experiences, and do not revise beliefs that are inconsistent with their delusions.[10] If delusions are acceptances, as I claim, then we do not need to posit a pathology of belief integration. It is simply in the nature of acceptances to not be adopted as beliefs.[11]

Thus: an acceptance that p is a disposition to use p, unvalidated, in Type 2 reasoning.[12] I can accept that p without believing it if I am disposed to reason from p

[10]To look ahead somewhat: because Bermúdez seeks to explain schizophrenic delusions by positing both a pathological experience and a pathology of reasoning, his is a *two-factor* theory of delusions. More on this to come.

[11]None of this is to say that inference from acceptance can have *no* effect direct on our beliefs. It is simply to say that unvalidated propositions on their own do not get taken into belief revision processes. After all, we often do assume p in order to form beliefs about what would follow from p. Given the uses of suppositional reasoning, there is very plausibly a mechanism that validates (and hence allows us integrate into our beliefs) conditional propositions of the form ([unvalidated proposition(s)] $\rightarrow$ [unvalidated proposition]) (Mercier and Sperber 2009, Recanati 2000).

[12]This is a necessary but not sufficient condition; it does not rule out states of pretense. I should comment on what it means to "be disposed to reason from p." Brian McLaughlin has pointed out to me that there is a sense in which I am disposed to reason from any proposition, because if you give me any proposition and ask me to reason from it, I can do so. I do not want to say that we accept all propositions, so this is not the disposition I have in mind. Some theorists (Cohen 1992, Frankish 2004) have attempted to navigate around this problem by saying that accepting a proposition involves setting a *policy* for oneself to reason from p, or *committing* oneself to reason from p. These words seem a little strong for the act that takes place when we form an assumption that p, and they invite further questions about what it means to "set a policy" that I think are unneccessary. I cannot explicate the disposition much more than I can by pointing to an intuitive difference between the state that one is in when one supposes that p and the state that one is in when one is prepared to suppose that p in certain circumstances. Both are dispositions to reason from p, but they are importantly different. For example, if a philosopher friend and I want to talk about some of the consequences of incompatibilism about free will, we might decide that for the rest of the night we will suppose that incompatibilism is true and see where conversation takes us. Then, throughout the night, we are in a standing state of supposition even when not actively reasoning about incompatibilism. This is a different state than I am in right now. I'm currently prepared to suppose that incompatibilism is true if I have good reason, but I am not supposing that incompatibilism is true. Similarly, there is a difference between pretending to be a duck and being prepared to pretend I am a duck in certain circumstances. Both are dispositions to quack, but they are importantly different.

without its consequences being integrated into my beliefs. I can believe that *p* without accepting it if *p* is always validated when it enters conscious reasoning, or if my belief is a tacit belief that I am not disposed to reason about consciously using Type 2 processes. I can be in the two states simultaneously if I tacitly believe *p*, but because I do not realize that I believe *p*, I accept that it is true and it enters into my reasoning unvalidated.

I next turn to a second feature of acceptance: its formation conditions.

6.2.2 Formation and Volition

One of the seminal papers in the literature on acceptances is Dennett's "How to Change Your Mind" (1978b). Dennett presents a notion of opinion which he contrasts with belief (as mentioned earlier, I'll pretend that he uses the word 'acceptance' rather than 'opinion'). Building off of previous work by de Sousa (1971) and Baier (1979), Dennett is animated by the thought that we need to posit a new attitude in order to make sense of the notion of "changing one's mind."

Belief formation and belief updating are sometimes presented as purely passive and involuntary processes, in which our store of beliefs is altered in response to incoming sensory evidence without intervention of the will. Hume (1739/2000, p.624) was an early figure to hold that belief formation is involuntary, and Williams (1973) has argued that this is not just an empirical fact, like the involuntariness of blushing. We cannot blush at will, but this is a contingent fact; it needn't have been so. Rather, according to Williams, the involuntariness of belief is a conceptual fact. It is in the nature of belief that it have truth as its aim, so we could not regard a state as a belief if we know that it has been formed without regard to its truth. Whether or not this argument is sound, the picture of belief as an involuntarily formed mental state is a popular one.

As well-regarded as it is, the picture runs counter to some very intuitive features

of belief. It appears that in ordinary life beliefs don't only occur to us—we are sometimes able to volitionally *make up our minds* about certain propositions. The Stoics called this sort of active judgment *sunkatathesis*. How can we explain this phenomenon and square it with the notion of beliefs passively updating in response to evidence? Dennett's solution is to claim that belief-updating really is a passive process; when we make up our minds about a certain topic, we form an acceptance, not a belief. Other acceptance theorists have followed in his stead.

However, many acceptance theorists claim that acceptances not only *can* be actively formed; they are *by definition* active, volitional, and under the control of the will.[13] Active commitment is how all acceptances are formed. It is true that paradigmatic acceptances are formed volitionally, as are assumptions, but this claim is usually simply stated outright rather than argued for. Part of the impetus here is clearly to distinguish beliefs from acceptances; declaring that one is necessarily active and the other necessarily passive serves well to make their differences apparent.

Not everyone agrees that acceptances must be formed volitionally. For instance, Tuomela (2000) holds that acceptances are only *typically* the result of volitional action.[14] I agree. It matters for my account of delusions that acceptances merely *can* be voluntarily formed (as I think that delusions are acceptances that are not formed voluntarily). It is telling that when many theorists speak of acceptances, their examples are often not good candidates for voluntary formation. For example, when discussing other features of acceptance, Dennett uses "Don Larson pitched the only perfect game in World Series history" as an example of an acceptance "par excellence" (1978b, p.306). This does not look like something that one would make up their mind about. It is not obvious why we should demand that acceptance must be voluntary. Better to say that we can, and often do, accept propositions voluntarily.

---

[13] E.g. (Cohen 1992, Engel 2000, Frankish 2004)

[14] "[I]t clearly seems possible that a person comes to accept that there is a tree in front of him without doing this at will or on purpose. As a matter of psychological fact, if it indeed is one, this acceptance is normally based on the person's causally induced belief that there is a tree in front of him. Given this, the acceptance involves reflection on ones belief (or beliefs) without yet being a mental action performed on purpose. For instance, the person does not in this case decide that there is a tree (rather than something else) in front of him, but finds himself to have this information" (Tuomela 2000, p.126).

6.2.3 Behavioral Effects and Verbal Effects

It is common to find theorists in the dual processing literature arguing for "the primacy of the implicit" (e.g. Reber (1993, chap.3)). According to these theorists, Type 1 reasoning is largely what drives us about in the world and controls our behavior. Experiments in social psychology conclusively show that many of our decisions are driven by nonconscious processes to which we have no introspective access. When individuals are asked why they have made certain decisions, they tend to confabulate explanations on the spot, and these confabulations have no relation to the underlying heuristics that determine their behavior (Nisbett and Wilson 1977). Given that acceptances do not play any role in Type 1 reasoning, what role can they play in the production of behavior? Moreover, explicating the notion of acceptance by appealing to assumptions and suppositions serves well to demonstrate their role in reasoning, but it serves to make them seem rather trivial or otiose. Normally, assumptions for the sake of argument do not have very sizable effects on our behavior: we assume that *p*, draw some derivations, then discharge the assumption and move on.

However, long-standing acceptances are not trivial or otiose; they have important behavioral consequences and play an important role in our lives. The examples of the murderer's son and the self-deceived philosopher and clergyman were meant to show that losing track of whether a state is a belief or an acceptance could have a profound influence on the course of a person's life. What behavioral and cognitive effects, then, do acceptances produce?

If you and I agree to suppose for the course of a night that incompatibilism about free will is true, our doing so will not only affect the thoughts that our minds turn to as we converse. It will also affect the flow of our conversation. Acceptances are often expressed *verbally*. In the previous chapter, I presented Stich's suggestion that there could be two sorts of databases in our heads, one responsible for directing our verbal reports, and another responsible for directing nonverbal behavior. This picture is close to what I have in mind. I do not, however, want to claim that all verbal behavior is generated from acceptance. Some of our utterances do express belief.

What I want to claim instead is that many times when we are called upon to speak, we are motivated to express not what we believe, but what we accept. Acceptances, like beliefs, interact with desires to produce certain behaviors and acts of reasoning. Accepting that *p* is not enough to cause us to reason from it; we must also *want* to reason from *p*. In a similar way, we often *want* to verbally express our acceptance, because doing so serves various useful functions. Thus, acceptances can conjoin with desire to produce behavior; this behavior is usually verbal because we are usually only motivated to speak our acceptances.

Delusions are not the only phenomenon in which saying and doing come apart. For many of these other phenomena, philosophers have posited acceptances in order to explain how there can be a mismatch between what one avows and how one behaves. For instance, Rey (1988) and Audi (1988) develop theories of self-deception of this sort. The self-deceived individual says one thing but behaves in another way because there is a mismatch between what he accepts and what he believes. Sperber (1996), Dennett (2007), and Rey (2007) explain insincere religious avowal by means of a belief/acceptance split. On weekdays, a person might loudly preach the existence of an eternal Hell, but go on to engage in all sorts of sinful behavior on weekends. Does this person really believe in Hell? Or do his actions reveal that he doesn't actually believe in Hell? Perhaps he accepts that Hell exists without believing it. A number of philosophers of science (Maher 1990) have argued that scientists bear a different attitude toward the content of their scientific theories than that of belief: they can be willing to assert a particular interpretation of quantum mechanics (say), even though they can have a low credence in it. Dennett (1978b) describes a smoker who says that he knows that smoking will give him cancer, even as he lights up a cigarette. An explanation here might be that he accepts he has a good chance of dying from his habit, but deep down, he really believes that his chances are low. There is an obvious analogy to draw between this case and the delusional patient who claims that his food is poisoned even as he eats it.[15]

[15]These examples should not be taken as explanations of *all* instances of saying/doing mismatch in self-deception, addiction to toxic substances, religious conviction, etc. For example, there are very

Language is deeply ingrained into many accounts of acceptance. Acceptance is often said to be "linguistically infected," and available to only natural language-using organisms (Dennett 1978b, Engel 2000). Some (e.g. Maher (1990)) have claimed that acceptance simply *is* a disposition to sincerely assent, where sincere assertion does not require belief (but these accounts ignore the important role that acceptance plays in reasoning). Sperber (1996, 2000) and Recanati (2000) claim that 'quasi-beliefs' are a type of acceptance. Quasi-belief is an attitude one bears toward sentences one is disposed to say that are meaningless or that have a content that the accepter does not fully understand. Sperber gives, as examples, the sentence "the Father, the Son, and the Holy Ghost are one" as spoken by a child at Sunday School, and the sentence "there are millions of Suns in the universe" as spoken by young Lisa who just heard her teacher say it. Lisa only knows the word 'Sun' as a name for the big glowing orb in the sky, so the claim is perplexing to her, but she nonetheless accepts it and is disposed to say it because she trusts her teacher. A five-year old might say "My daddy is a doctor" without having the slightest idea what it means for someone to be a doctor (Dennett 1969, p.125); does he really have the belief that his daddy is a doctor?[16] A little reflection reveals that many assertions are of things that the speaker does not fully comprehend.[17]

---

plausibly cases of self-deception that do not conform to an acceptance-based form of explanation. On Mele's (2001) theory of self-deception, a person who is self-deceived that p suspects that $\neg p$ might turn out to be the case if evidence were collected, but has strong prudential reason for believing p, so he or she orients attention away from potential sources of evidence that $\neg p$. This is a perfectly comprehensible description of what might take place in certain peoples' cognition, and it would produce behavior that we would naturally call self-deception. I do not wish to impugn it. However, I do not think that this should be taken as an explanation of all behavior that we would call self-deceived behavior. There is a bit of a sleight-of-hand involved in presenting *the* puzzle of self deception—as if all self-deceived behavior must be the product of a single cognitive phenomenon for which we can discover *the* explanation. Similarly, some nicotine addicts might believe that their chances of dying are low and casually say "yeah, I know it'll kill me" when they are challenged upon lighting up. Others might be very worried whenever they find themselves smoking because they have genuinely internalized the dangers. An acceptance-based explanation might serve to explain the former case but not the latter.

[16] Dennett uses this example to show that belief comes in degrees; he does not consider that it might well be better explained by his account of acceptance.

[17] On this account, acceptance is an attitude toward a sentence in some relevantly English-like language (possibly syntactically disambiguated). I have occasionally spoken of "acceptance-that-*p*" where p is a proposition; this should be parsed as "acceptance-that-*s*, where *s* is a sentence that expresses proposition *p*." Sometimes it is claimed that delusions have "impossible" contents (Jaspers 1963). Although I think most alleged cases of impossible contents are merely unlikely, if there are any delusions

Why would one be disposed to speak one's acceptances? There are a number of reasons; language has a variety of functions. It's easy to be tempted by a myth about language that goes something like the following: the function of language is to communicate information, and it exists to help us share knowledge with one another. By communicating my beliefs with you, I add to a collective store of knowledge that aids in the construction of society. When I lie to you or tell you things that I don't really believe, I contravene norms that govern the proper use of language. From the perspective of this myth, it is difficult to see what value there would be in verbally communicating to you things that I do not actually believe. However, language has many more functions than this simple myth lets on, and not all of these involve communicating belief. Verbally expressing our acceptances, and forming acceptances at all, has various forms of important prudential value. Forming a long-standing acceptance and being disposed to verbally assent to it has both a cognitive function and social function.

Cognitively: we often volitionally commit ourselves to accept propositions in order to change our beliefs and other mental states: it is a roundabout way of changing beliefs that are not directly subject to our control. Lisa might accept "there are millions of Suns in the universe" in order to facilitate conversation that will help her eventually understand what it means. However, coming to understand a cryptic acceptance is not the only way that an acceptance can enter into our belief store. For example, I might discover in myself a host of implicit racist or sexist beliefs that I do not want to have, so I plant my flag and accept the equality of all. There are a few ways this can influence me. Firstly, it redirects my attention to arguments that will hopefully influence my beliefs. The pastor quoted at the end of my Unbelieving Clergyman vignette hoped to discover a compelling argument for God by taking up the acceptance. Speaking our acceptances out loud affords us the ability to collectively reason with another person, as happens when you and I assume something

---

with impossible contents, an acceptance-based account could accommodate them. It is also possible that some of truly bizarre utterances produced through formal thought disorder ("word salad") are delusions without meaningful content. I doubt they are, as I doubt that delusional individuals reason from them, but I admit the possibility.

for the sake of argument and then have a conversation. Secondly, it causes me to be aware of my thoughts and actions that might contradict the consequences of my anti-racist acceptances. Thirdly, it is possible that by saying something enough, I will eventually come to believe it through various non-rational biases. For example, experiments from Gilbert et al. (1993) suggest that we are apt to raise our credence in what we read even if we are explicitly told that what we are reading is false. Exploiting this bias, simply saying things you believe are false can plausibly help get you to believe that they are true. Fourthly, I might confuse my expressions of acceptance for expressions of belief—something that is likely for the folk to do, given the fact that the word 'belief' covers both attitudes and they have only a nascent understanding of the distinction—I will form a metarepresentational belief that I believe that all races and sexes are morally equal. If I take myself to be fairly reliable at forming beliefs, I should think that I believing p provides evidence for p, and I should raise my credence in p. The trickle-down cognitive effects of acceptance can also affect attitudes other than belief as well. A person might proclaim love for her partner in the hope that she will eventually come to feel it.

Socially: the benefits of accepting non-sexist and non-racist propositions, and avowing them, should be fairly obvious. Expressing one's acceptance of p is a speech act that publicly expresses one's commitment to a cause and pledges one's allegiance. The metaphor of "flag-planting" is helpful to describe this function. One can plant one's flag on all sorts of subjects: from the sacred (the truth or falsity of various religious doctrines) to the profane (the worth of a particular sports team or political tenet). For instance, we might commit ourselves in order to display public allegiance to a cause with which we want to be associated. This comports well with the commendatory function of the use of the word 'belief' in ordinary folk English; the word is often used to avow faith and commitment, rather to express the result of epistemic deliberation. It is not hard to imagine why the person would be motivated to declare his or her acceptance. It lets the world know the sorts of claims that the speaker is willing to stand by and the sort of person the speaker has pledged to be. We might

even declare our acceptances to ourselves in order to discover what our commitments truly are.

Expressing allegiance is not a necessary feature or function of acceptance, but it is one prominent use to which acceptance is put. I think it is likely that many ordinary public utterances are intended to be expressions of acceptance rather than expressions of belief. We might call these speech acts 'endorsements' to distinguish them from assertions. It's easy to fall into the trap of thinking of conversation as the mere swapping of assertions, but this is an idealized an unrealistic picture; we assert and endorse (and produce many other speech acts besides), and it is often not clear to us which mental state we are expressing at any one time.

The social and cognitive functions can both be served by a single acceptance. We often defer to authority figures and accept what they say. Reasoning from their premises and avowing their premises in public can have the dual function of manifesting one's commitment to a figure's authority, as well as beginning a step into fully understanding the premise and integrating it into one's beliefs.

However, it is too strong to say that acceptances only affect verbal behavior, even though many have made this claim. According to Dennett (1978b), acceptances control what we say, but are otherwise mostly inefficacious. Frankish (2004, p.75) points out that this picture cannot be exactly right: acceptances cannot be as causally circumscribed as Dennett would have it. We are often motivated to verbally express our acceptances because doing so serves certain functions, but other behaviors can also serve these functions. In such cases, we can be motivated to act on our acceptances in certain non-verbal ways. For example, if you and I decide over dinner to suppose that incompatibilism is true, we might not just talk about it; we might also jot down diagrams on a napkin. Doing so instrumentally helps us reason. A person who accepts that God exists in order to not be exiled from her social group might not just publicly declare her allegiance; a desire to make this acceptance known can cause her to go to church. Acceptances can therefore result in some non-verbal behaviors, but behavioral manifestations of acceptance are normally verbal.

Thus concludes my characterization of acceptance. An acceptance that *p* is a disposition to draw conclusions from a sentence *s* that expresses *p* in Type 2 reasoning; it is typically though not necessarily volitionally formed; the acceptance manifests itself in verbal behavior because we are often motivated to verbally express our acceptances, and it is often used to indirectly alter other mental states and to publicly affirm commitment or allegiance.[18] All of the description above should have made it reasonably clear how acceptance provides a good model for delusion, but I will make the case explicit in the next section.

6.3 Delusion as Acceptance

On the model I propose, delusions are pathologically-formed acceptances that conflict with the deluded individual's beliefs.[19] To be deluded that one's wife has been replaced with an imposter is to accept that one's wife has been replaced with an imposter. Doxasticists such as Bortolotti have argued against non-doxasticism on the basis that delusions share their allegedly bizarre features with nonpathological mental states, but this is in fact a point in favor of an acceptance-based model. These non-pathological mental states are themselves not beliefs.

I consider below the features of delusion that are resistant to explanation in a doxasticist model, and show how they are, in fact, paradigmatic features of acceptance.

6.3.1 Explaining Double-Bookkeeping

'Double-bookkeeping' refers to the collection of phenomena that arise because subjects appear to keep two "books" of representations in their heads—one of "the real

[18]The reader who is familiar with Gendler's "aliefs" (2008a, 2008b) may be struck by similarities between her proposal and the one developed here. We both agree that the ordinary term 'belief' picks out two sorts of mental state that ought to be distinguished, though she uses 'belief' to denote the state that I would call 'acceptance'. Gendler does not use the Type 1/Type 2 distinction to illuminate her view, though others such as Kriegel (2012) have done so. I won't explicate her views on the matter, but the differences are numerous enough to keep our views distinct. (For instance, on her view, the contents of aliefs are not propositional, whereas I do think beliefs are propositional.)

[19]Frankish (2009) also suggests that an acceptance-based model can be used to explain delusions, though his conception of acceptances differs from mine.

world" and one of "the delusional world"—each of which controls behavior and cognition in a different way. The distinction between belief and acceptance is also often described in this way (or, more often, in terms of there being two distinct boxes of representations in the head: a "belief box" and an "acceptance box."). Double-bookkeeping is excellent evidence that delusional subjects bear two different attitudes toward a proposition and its denial.

When double-bookkeeping was first presented, I mentioned three possible phenomena that the term could refer to:

BEHAVIOR GUIDANCE: Delusions are behaviorally circumscribed, but this does not mean that patients lie there inert; instead, their actions seem to rely on a *different* knowledge store.

METACOGNITION: Delusional subjects seem to have a *metacognitive knowledge* that there is a division of some sort in what they take to be true.

PHENOMENAL EXPERIENCE: Delusional subjects have a phenomenal experience of "two worlds" overlaid upon one another.

The first in the list is the easiest for an acceptance-based account to accommodate. Delusional individuals seem to behave as if they believe $\neg p$ even when they are deluded that p, and this is because delusional individuals *do* believe $\neg p$—it is written into the "belief book" that controls much of their behavior.

The metacognitive feature of double-bookkeeping reveals that delusional subjects can tacitly acknowledge that there is something different in their attitudes toward the content of their delusion and the content of their beliefs, but are unable to articulate what it is. We can often recognize that we accept something that we do not believe. I doubt that we have perfect introspective access to our attitudes. When I am speaking, I do not always know whether I am expressing what I believe or what I merely accept. My beliefs and my acceptances are revealed to me as I see how I react in various scenarios; I might be surprised to discover that I do not believe what I have been alleging for many years. Delusional individuals are constantly buffeted

with information that there is a split between their verbal behavior and their nonverbal behavior, so they are aware that there is some sort of cognitive split, but lack the resources to describe it. Their difficulty in describing their situation is not surprising. Folk psychology is underdeveloped and the word 'belief' is used to refer to both states; even non-deluded individuals would be hard-pressed to describe the difference between their beliefs and acceptances. A regular person-on-the-street might know that there is something importantly different between her attitude toward the claim that God exists and the claim that Lima is the capital of Peru, but be unable to say exactly what this difference amounts to.[20]

Finally, there is the fact that delusional subjects have a certain sort of split phenomenal experience. The account of acceptances that I have described does not make many claims about the phenomenology of acceptance. Type 2 reasoning is typically taken to be conscious, so when an acceptance is actually manifested in reasoning, one will have the experience of reasoning with it. Aside from that, there are many open questions here about how an acceptance/belief mismatch will affect one's experience. Still, it would not be surprising if the experience were reported as being of "two worlds." Pretense is not acceptance, but it is a near neighbor of acceptance, and there is a sense in which pretending that p when you know that $\neg p$ gives a kind of experience of one world overlaid on another. The similarities and differences between the experience of pretense, the experience of an acceptance/belief mismatch, and the experience of delusion, is a topic I leave for future researchers.

---

[20] I am curious to know whether a delusional individual who understood and internalized a theory of acceptances would be willing to describe their delusions as acceptances. I should note that I do not think that teaching delusional individuals that their delusions are "mere" acceptances will be clinically useful. Recognizing that an attitude is an acceptance and not a belief will not in itself cause one to discharge it. For example, I am aware that many of my more progressive anti-racist and anti-sexist attitudes are acceptances, but I see the value in them, so I am not motivated to get rid of them. Delusional individuals don't have the same prudential motivation to hang on to their acceptances, but they do have a pathology that caused them to form the acceptance in the first place, which will likely block them from discharging the acceptance (as I will describe shortly, this pathology is a pathological cognitive feeling). In any event, I do not have a theory about the conditions that cause one to lose acceptances. I am not sure that it is always a willed action or that willed action always suffices.

In any event, the core idea underlying the phenomenon of double-bookkeeping—that behavior appears to be controlled by two different stores of information—is easily explained by the fact that, in an acceptance-based theory of delusion, behavior is controlled by two different stores of information.

6.3.2 Explaining Circumscription

Circumscription is also easily explained if delusions are acceptances. Delusions are *theoretically circumscribed* because unvalidated acceptances are theoretically circumscribed. They do not generate new beliefs after being manifested in inference. Moreover, an acceptance-based account can explain how delusional individuals can spin off long confabulations when prodded. The individuals are fully able to reason using their delusional premises and give explanations for how their delusion could possibly be true, yet because these confabulations are themselves unvalidated, they are not integrated into the belief store and are manifested only verbally.

Delusions are *behaviorally circumscribed* because acceptances are behaviorally circumscribed. Accepting that p does not yield the behavior that one would exhibit if one believed that p. It primarily produces verbal behavior related to saying that p.

However, delusions are not completely behaviorally circumscribed, and delusional subjects do sometimes act on their delusions in various ways. How is this possible if delusions are acceptances? There are two possibilities. Firstly, it could be the case that delusional content sometimes comes to be believed. I earlier mentioned that one of the reasons to form an acceptance is to try to get its content to trickle down into belief. Repeated exposure to hearing oneself report a delusion could gradually change one's credences and weaken one's cognitive defenses, eventually resulting in a full-blown belief. If this were to happen, the delusional subject's actions would be those of a true believer; he would no longer exhibit some but not all of the characteristic features of belief. This might occur in cases in which a person commits very violent acts such as murdering the suspected imposter, but I think that it is rather rare if it ever happens at all. We don't see delusional subjects suddenly act completely like true believers: circumscription, double-bookkeeping, and

unresponsiveness to evidence remain.

Secondly, acceptances do not instigate *only* verbal behavior; nonverbal behavior can arise when the person has motivation to act on their acceptances. It could be the case that delusional subjects are motivated to act on their acceptances in certain ways for their own prudential reasons. I think this is the likelier of the two explanations. However, given the huge variety of reasons that delusional subjects could have for acting on their acceptances, and given that no one has yet formulated any general principles for when delusions results in action, I doubt that a general theory of delusional behavior will be forthcoming. Only case-by-case studies and close analyses will be able to tell us when and why delusional individuals act as they do.

Let's consider a case. A compelling example was recently related to me by woman with schizophrenia, Sophie.[21] At one point in her history, a therapist who Sophie was fond of was reassigned. The two would not be able to see each other any more. Sophie became extremely upset and distraught, and began to feel betrayed. Eventually, during a psychotic episode, she formed a Capgras-like delusion: the therapist had been killed and replaced with an imposter. She e-mailed this news to all of the therapist's colleagues in her psychology department. Looking back on the situation, Sophie claims that her delusion was a defense mechanism intended to soothe the sting of betrayal. In Sophie's delusion, her therapist didn't actually betray her; her therapist was killed. When asked why she sent the e-mail to the department, Sophie responded that doing so was a form of public commitment. Publicly committing herself to the truth of the delusional proposition made it seem, in some sense, more real.

In this case, Sophie's delusion was motivated. Sophie did not want to believe that her therapist had willingly left her. So, she formed the acceptance that the therapist who was avoiding her was not *her* therapist. I think the best interpretation of her claim that she wanted the delusion to feel "more real" was that she recognized that she didn't actually believe the delusion; e-mailing the department was an attempted

[21] 'Sophie' is a false name used for the sake of anonymity.

way of getting herself to fully integrate the acceptance into her beliefs. Why would e-mailing the department make the claim seem more real? This is a difficult question, and close investigation of this particular case would be required to really get to the bottom of this, but here are a couple of hypotheses. Firstly, doing something as drastic as e-mailing a whole department is something that someone rational would only do if they were absolutely sure that the imposter hypothesis was correct, so her actions are an attempt to trick herself (so to speak) into believing that she believes it. Secondly, the public nature of her avowal is important. It is easy to feel solipsistic in one's delusions. E-mailing everyone was Sophie's way of planting her flag and shouting to others that her allegiances lie with whoever else agrees that her therapist had been killed. It manifests something like a desire for others to get involved, and hope that others will rally to her cause and help convince her that she was right after all.

Delusions are also often *affectively circumscribed*. A delusional individual might well be strangely unperturbed and not anxious about the situation he claims to be in. This feature also comports well with acceptance; to accept that p does not require having the affective responses one would associate with a belief that p. However, both delusions and acceptances *can* result in strong affect. Delusions can be scary and cause anxiety, but so too can acceptances. Accepting a proposition that describes a frightening or anxiety-provoking state of affairs can cause one to imagine that scenario, and simply imagining an anxiety-provoking scenario provokes anxiety.

6.3.3 Unresponsiveness to Evidence

Finally, delusions are unresponsive to evidence. Once again, acceptances share this feature. Acceptances are not formed on the basis of evidence, and one cannot argue someone out of an acceptance just by showing them that the acceptance is inconsistent with propositions they believe.

However, we are still left with a question about how delusional acceptances are formed, if acceptances are not formed on the basis of evidence. Acceptances are typically formed through volitional action for prudential reasons. Although some delusions are arguably motivated and are products of volition, such as those that are

erected as a kind of psychic defense against whatever cognitive catastrophe would occur otherwise, many are not motivated at all. Most instances of the Capgras delusion are unmotivated (unlike the delusion of Sophie's which was just mentioned). The typical Capgras patient accepts that his wife is replaced with an imposter without being motivated to accept it and without deciding to accept it. Some account is required to explain how the delusional acceptance got lodged in his head.

The delusion, I claim, was forced upon him by a pathological cognitive feeling. I explain how this can be so in the next section.

6.4 Cognitive Feelings

In the first chapter, I claimed that unresponsiveness to evidence does not strongly motivate non-doxasticism, and this is because doxasticists have made some measure of progress on this front. However, it's still something that the non-doxasticist needs to explain. How does the bizarre acceptance get implanted in the first place, and why is it so difficult to dislodge, even when it causes distress? Here, I'll present what I consider to be the best theory of delusional belief acquisition, and show how it is easily and profitably adapted to a theory of delusional acceptance formation.

Among theories of delusion formation are one-factor and two-factor theories. One-factor theorists such as Maher (1974, 1999) claim that a theory of delusion formation need only posit a single pathology: an anomalous experience. Two-factor theorists (such as that of Davies et al. (2001)) claim that an anomalous experience can explain why a delusional hypothesis is considered, but not why it is adopted. A second deficit—a pathology in reasoning—must be posited. The main problem with two-factor theories is that there exist delusional individuals who do not exhibit any deficit in reasoning about topics unrelated to their delusion. This is particularly acute in monothematic delusions such as Capgras, in which a subject has delusions on only a single particular theme, and otherwise approximates normal standards of rationality. If a subject suffered an inability to reject the contents of experience—what Davies et al. call "a failure to inhibit a pre-potent doxastic response"—one

would expect that the subject would be taken in by all sorts of sensory and optical illusions. But this does not happen.

Here is a dilemma (McLaughlin 2009). One-factor theorists have difficulty explaining why delusional subjects irrationally adopt and hold onto their belief; two-factor theorists have difficulty explaining why the irrationality is apparent in only certain domains. This is where the debate currently stands. The debate is easily translatable to an acceptance-based theory. A one-factor acceptance-based theory claims that a delusional acceptance is formed in response to an aberrant experience; a two-factor account claims that a delusional acceptance is formed in response to a pathological experience and maintained because of a deficit in reasoning.

Notice firstly that there seems to be no reason to posit a second deficit on an acceptance-based theory. A second deficit is posited to explain why rationality does not prevent the delusion from being fixed as a belief, but on an acceptance-based theory, delusions are not fixed as beliefs. Acceptances can be (but are not necessarily) formed for pragmatic purposes or by voluntary decision. If an acceptance is formed for epistemically irrational reasons, it does not inspire as much mystery as does a belief that is formed for epistemically irrational reasons. Similarly, it is not mysterious why acceptances aren't rejected for clashing with other beliefs. Many of the things we accept clash with our implicit beliefs. This does not explain how the delusional acceptance is implanted, but it does suggest that the non-doxasticist should adopt a one-factor theory.

One way for the one-factor theorist to break the standstill would be to identify a subset of experiences that automatically initiate belief formation. Suppose that there are certain experiences which are taken up unquestioningly by a belief-forming mechanism. The content of the experience is directly believed. A one-factor theorist could use these sorts of experiences to explain why delusional beliefs are acquired. This is the position taken by McLaughlin (2009). On his one-factor theory, delusions are formed in response to powerful *cognitive feelings*, such as the feeling of unfamiliarity, or the feeling of importance. Ratcliffe (2005) presents a similar feelings-based theory of delusion.

It is thought that there are two visual pathways involved in face recognition. The ventral route of the visual system is responsible for conscious, overt facial recognition, and the ventral limbic structure projecting to the amygdala is responsible for a covert, affective response. In many Capgras patients, a brain lesion has caused damage to the covert, affective pathway, but not overt, conscious pathway. Thus, there is conscious recognition combined with a lack of affect, and this combination generates a powerful feeling of unfamiliarity. The loved one "looks right" but "feels wrong" (Langdon and Coltheart 2000, p.187). He or she feels unfamiliar and alien; these feelings are so overbearing that they push aside the fact that the person visually appears unchanged.

Other delusions also involve powerful feelings. The schizophrenic experiencing delusions of reference feels that ordinary objects are extremely important to her in some way. The Cotard patient feels emptiness and insignificance. What are we to make of these experiences? We can all grasp what is meant by a feeling of understanding, a feeling of safety, a feeling of importance, or a feeling of familiarity.[22] Think of how familiar your neighborhood feels to you now compared to how it felt when you first moved in. These sorts of feelings are ubiquitous in everyday life, and are also very naturally deployed in discussions of delusion formation. Yet, they are not easily captured by the mental states and attitudes typically discussed by philosophers. They are not beliefs or desires, nor are they visual or somatic perceptions, nor are they emotions or moods.

---

[22] These are some of the attitudes that have been mentioned in print; this literature is in its infancy, so a complete taxonomy and characterization of cognitive feelings is a matter for future empirical work. Schwarz and Clore (1996, p.386) claim that cognitive feelings include surprise, amazement, and feelings of familiarity, and they are so-called because they "inform us about knowledge states." By this, they mean that such states carry information about our cognition: a feeling of familiarity carries information about whether we have cognized someone before. This would exclude a number of states that are intuitively like cognitive feelings, like feelings of religiosity or safety or importance, and would be of less use in explaining a broad number of delusions, so I prefer to think of cognitive feelings more broadly. It is the case, however, that most of the cognitive feelings mentioned in the literature have an *egocentric* component to them—they imply something about the person experiencing the feeling, if not about his or her cognition. A person isn't familiar *simpliciter*, they are familiar to *me*; a place isn't safe *simpliciter*, it is safe for *me*. This might be a characteristic component of the content of cognitive feelings.

According to McLaughlin, cognitive feelings are intentional states with propositional cognitive content: an individual feels *that* a person is familiar.[23] The cognitive psychologists Clore and Gasper (2000) are the most prominent psychologists investigating cognitive feelings; their experimental evidence has led them to propose "a feelings-based hypothesis, which says that cognitive [...] feelings, despite the fact that they are self-produced, may be experienced as internal evidence for beliefs that rivals the power of external evidence from the environment" (2000, p.26). According to McLaughlin, this explains the acquisition of delusional beliefs. Our belief-forming mechanisms respond to feelings as evidence; particularly powerful feelings are taken as particularly powerful evidence. In normal individuals, feelings are reliable. By and large, what feels familiar *is* familiar. In cases where the two diverge, we experience an illusory feeling. Delusional patients experience these sorts of illusory feelings, and they are powerful enough to overwhelm counterevidence and immediately implant their contents into the patients' stores of beliefs.

I find this story compelling. It circumvents the normal problems with one-factor theories and does justice to the fact that most theorists do talk about feelings in their explications of delusion. However, as a doxasticist theory, it does not explain double-bookkeeping or circumscription. McLaughlin should expect delusional subjects to behave as if they really did believe their delusions; because they do not, we should try to adapt the theory to an acceptance-based account. The natural way would be to claim that feelings do not influence beliefs; they influence acceptances. A powerful feeling that p tends to leads us to immediately accept p, which can happen without our coming to believe that p. So, for instance, a powerful feeling of religiosity will immediately generate a religious acceptance; a powerful feeling of superiority or invincibility will lead to narcissistic acceptances. If a person feels something strongly, they become committed to reasoning from it. This is the account of involuntary acceptance formation that I endorse.

---

[23] He contrasts cognitive feelings with background feelings; the distinction roughly mirrors that of moods and emotions. I will not discuss background feelings.

6.4.1 Acceptances and Cognitive Feelings

One of the few places in the psychological literature where feelings make an appearance is in the literature on dual process theory. A dual process theorist faces a problem of predicting when Type 2 processing will be invoked in reasoning. What conditions prompt a new proposition being taken up by Type 2 reasoning? One explanation on the table invokes a "feeling of rightness" (Thompson 2009). Feelings of certainty and uncertainty determine whether we deliberate on a proposition; the strength of a feeling of rightness determines the probability that Type 2 reasoning will be engaged to work on a problem or to rethink a decision. When we strongly feel that a proposition might not be right—when we feel uncertain—we then take up the proposition in reasoning.

In this case, a feeling with the content "I am uncertain about whether *p*" does not appear to generate beliefs. Rather, it modulates our reasoning about *p*, and this appears to be the primary function of the feeling. I claim that all cognitive feelings serve this function.

Consider familiarity. McLaughlin might be right that powerful feelings, when they appear, are reliable indicators of their contents. When we feel that someone is unfamiliar, they often are. However, the converse is not the case: someone's being familiar or unfamiliar does not necessarily generate the corresponding feeling in us (or at least, the strength of one's feeling does not correspond with degree of familiarity). Every stranger that I encounter is unfamiliar, yet I don't experience powerful feelings of unfamiliarity when meeting every stranger. Whatever mechanism generates feelings is like a geiger counter that goes off only when there is radiation nearby, but does not go off every time there is radiation nearby.

Others have commented on how it is difficult to determine why we feel familiarity or unfamiliarity in some cases but not others. For example: *prosopagnosia* is typically said to be the converse of Capgras Syndrome: there is damage to the overt visual route but not the covert visual route. Prosopagnosics are unable to consciously recognize faces—they are unable to tell you who they are looking at—yet they still

have the appropriate affective response when looking at loved ones, as measured by skin conductance response (Bauer 1984). However, while the lack of positive affect in the Capgras delusion (as measured by skin conductance response) is experienced as a powerful sense of unfamiliarity, the presence of positive affect in prosopagnosia does not contribute to any strong sense of familiarity. There is a phenomenological asymmetry (Young 2007).

When do we experience feelings of unfamiliarity, if not simply when among the unfamiliar? It is when we are among the unfamiliar and this fact demands our attention. We feel strong feelings of unfamiliarity when we are lost, or when in a dangerous neighborhood, or when confronted with someone who greets us by name but who we cannot remember. Contrariwise, strong feelings of familiarity occur when we recognize someone but cannot place their name, but not when we run into a colleague at work who we see day-in-day-out. Strong feelings often tell us that something is potentially amiss, and so they force us to ruminate on that fact, and direct our deliberative reasoning toward it.

Type 2 processes are very often called upon to rationalize why a certain judgment is correct (Evans 1996). When this occurs, a proposition is accepted as true, and the person searches for justification. He or she *assumes* that the proposition is true in order to make a case for it, as a lawyer does in order to make a case for a client. Thus, feelings exist to instill in us an acceptance when we need one most—a kind of pretend, temporary belief that directs our attention to resolving something in the world that is mysterious. Assumptions are normally temporary; acceptances based upon cognitive feelings normally wane with the passing of the feeling. However, if the feeling is continuous and powerful, the proposition never gets discharged, and it is used as a premise in reasoning for a long time. Déjà vu gives us a feeling of familiarity which causes us to stop and say, "wait, I've experienced this before." Once we realize that there is no way that events could be repeating themselves, the feeling subsides and we discharge the assumption that the events taking place are familair. However, it is well-known that in cases of chronic déjà, the sufferer displays behavior that looks delusion-like. When a patient is taken to the doctor for the first time, he might say,

"I'm not sure why you've brought me here again; it didn't work last time."

A virtue of this acceptance-and-feeling-based model is that it explains psychological pathologies that are not delusions, but that are similar to delusions. Jennifer Nagel (2012) has drawn attention to the phenomenon of spider phobia. Arachnophobes have responses to spiders that, like delusions, are "often extremely resistant to corrective verbal information" (Baeyens et al. 1992, p.134), and that they recognize are irrational (Mayer et al. 2000). However, experimental measures of implicit attitude (which measure the response times of individuals asked to match pictures of spiders with either positive or negative words) do not reveal any differences in the implicit attitudes or implicit beliefs of arachnophobes and individuals unafraid of spiders (de Jong *et al.* 2002). Arachnophobes have planted their flag in the awfulness of spiders ('Spiders are horrible!'), but the implicit attitude test reveals that their underlying beliefs do not match.

The authors of one of the studies hypothesize that "the nonfearful individual is the one who can override [an] automatic negative stereotype, whereas the phobic individual is the one who does not attempt or is not able to control it" (de Jong et al. 2003, p.540). Note that this explanation leads to something like the same dilemma that confronted us when deciding between one-factor and two-factor theories. Does the arachnophobe have a *general* tendency to be swayed by automatic negative stereotypes? We can avoid this conclusion by appealing to what seems to me a plausible explanation: arachnophobes simply feel a much stronger sense of revulsion at spiders. Spiders cause a powerful feeling of creepiness in the phobic individuals that immediately implants 'Spiders are horrible' (or something in the neighborhood) into the patient's acceptances. All of this suggests a similar experiment one could do with delusional patients. The Capgras patient will feel that a picture of his spouse is unfamiliar and will claim that it is not his spouse. What will an implicit attitude test reveal? I predict that he will exhibit positive attitudes. He will be quicker to match positive words with a photo of his spouse than negative words.

6.5 The Final Model

Here is how the final model might be applied to an instance of Capgras Syndrome. When confronted with her spouse, the Capgras patient experiences a powerful feeling of unfamiliarity, typically due to a brain lesion. The feeling that the person before her is unfamiliar causes her to accept that the person before her is unfamiliar. She is now disposed to use this proposition as a premise in reasoning in a wide variety of contexts. The continued presence of the spouse and revisited memories of him cause continuing feelings of unfamiliarity, and this ensures that her acceptance never wanes away. However, she never comes to believe it, as it is unvalidated. So, nothing with delusional content comes to be fixed as belief. Because beliefs are largely responsible for her behavior, and because she continues to believe that the man is her husband, her behavior is circumscribed, and she continues to live with the purported imposter. Nonetheless, she is often driven to verbally express what she accepts as true, and this results in her stubborn expressions of delusion.

Similar explanations can be invoked for other delusions. A person suffering from the Cotard delusion might pathologically feel intense feelings of emptiness or remove; these straightaway form in him the acceptance that he is empty or dead. A schizophrenic delusion of reference might be brought about by feelings that objects nearby are personally significant in some way. Delusions of catastrophe are the result of feelings of impending doom.

6.6 Conclusion

Clinical delusions are difficult to explain in a belief-desire framework that is sparsely populated with only a few broad sorts of mental state categories. The model I propose suggests we will find success if we admit acceptances and cognitive feelings into our models of cognitive architecture. *Double-bookkeeping* can be explained in terms of the belief/acceptance distinction. *Circumscription* can be explained by the insulation of acceptance from belief, and the fact that we act on our beliefs, but verbally express our acceptances. Finally, *unresponsiveness to evidence* is explained by

the fact that delusional individuals suffer from pathological cognitive feelings, and powerful cognitive feelings normally cause one to immediately accept the content of the feeling.

The model I have offered is ambitious. It presents not only a revisionary theory of delusion, but a revisionary theory of normal cognition as well. Given the scope of the project, I am not sure that all details of the proposal are correct; I *am* sure that the theory is nearly correct, and that many delusions are belief-like acceptance states that are generated by certain sorts of feeling. Research is ongoing, and I am optimistic about future prospects and applications. Although I hope that my readers have found good evidence for believing this account, I would nonetheless be delighted if they have found reason to accept it.

Bibliography

Addis, D., Pan, L., Vu, M., Laiser, N., and Schacter, D. (2009). Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. *Neuropsychologia*, 47(11):2222–2238.

Aimola Davies, A. and Davies, M. (2009). Explaining pathologies of belief. In Broome, M. and Bortolotti, L., editors, *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, pages 285–326. Oxford University Press.

Atran, S. (2002). *In Gods We Trust: The Evolutionary Landscape of Religion*. Oxford University Press.

Audi, R. (1988). Self-deception, rationalization, and reasons for acting. In McLaughlin, B. P. and Rorty, A. O., editors, *Perspectives on Self-Deception*. University of California Press.

Austin, J. L. (1962). *Sense and Sensibilia*. Oxford University Press.

Baeyens, F., Eelen, P., Crombez, G., and Van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy*, 30(2):133–142.

Baier, A. (1979). Mind and change of mind. *Midwest Studies in Philosophy*, 4(1):157–176.

Bain, A. (1859). *The Emotions and the Will*. D. Appelton.

Bauer, R. M. (1984). Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the guilty knowledge test. *Neuropsychologia*, 22(4):457–469.

Bayne, T. and Hattiangadi, A. (manuscript). Belief and its bedfellows.

Bayne, T. J. and Pacherie, E. (2004). Bottom-up or top-down: Campbell's rationalist account of monothematic delusions. *Philosophy, Psychiatry and Psychology*, 11(1):1–11.

Bayne, T. J. and Pacherie, E. (2005). In defence of the doxastic conception of delusions. *Mind and Language*, 20(2):163–88.

Bentall, R. P. (2003). *Madness Explained: Psychosis and Human Nature*. Penguin Books.

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

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

Bleuler, E. (1950). *Dementia praecox; or, The group of schizophrenias*. Monograph series on schizophrenia. International Universities Press.

Block, N. (1995). Some concepts of consciousness. *Sciences*, 18:2.

Blount, G. (1986). Dangerousness of patients with capgras syndrome. *Nebraska Medical Journal*, 71(207).

Bortolotti, L. (2009a). Delusion. In *Stanford Encyclopedia of Philosophy*.

Bortolotti, L. (2009b). *Delusions and Other Irrational Beliefs*. Oxford University Press.

Bortolotti, L. (2011a). Continuing commentary: Shaking the bedrock. *Philosophy, Psychiatry, and Psychology*, 18(1):77–87.

Bortolotti, L. (2011b). In defence of modest doxasticism about delusions. *Neuroethics*, 5(1):39–53.

Bortolotti, L. and Broome, M. (2008). Delusional beliefs and reason giving. *Philosophical Psychology*, 21(3):1–21.

Boyer, P. (1994). *The Naturalness of Religious Ideas: A Cognitive Theory of Religion*. University of California Press.

Braddon-Mitchell, D. and Nola, R. (2009). Introducing the canberra plan. In Braddon-Mitchell, D. and Nola, R., editors, *Conceptual Analysis and Philosophical Naturalism*, pages 1–20. Mit Press.

Braithwaite, R. B. (1932). The nature of believing. *Proceedings of the Aristotelian Society*, 33:129–146.

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Center for the Study of Language and Information.

Buchanan, A., Reed, A., Wessely, S., Garety, P., Taylor, P., Grubin, D., and Dunn, G. (1993). Acting on delusions. ii: The phenomenological correlates of acting on delusions. *The British Journal of Psychiatry*, 163(1):77–81.

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

Carpenter, P. K. (1989). Descriptions of schizophrenia in the psychiatry of georgian britain: John haslam and james tilly matthews. *Comprehensive psychiatry*, 30(4):332–338.

Chalmers, D. (2009). Ontological anti-realism. In *Metametaphysics: New Essays in the Foundations of Ontology*, pages 77–129. Oxford University Press.

Chalmers, D. J., Manley, D., and Wasserman, R., editors (2009). *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University Press.

Chan, T. (forthcoming). Introduction: Aiming at truth. In Chan, T., editor, *The Aim of Belief*. Oxford University Press.

Cherniak, C. (1986). *Minimal Rationality*. MIT Press.

Clore, G. (1992). Cognitive phenomenology: Feelings and the construction of judgment. *The construction of social judgments*, pages 133–163.

Clore, G. and Gasper, K. (2000). Feeling is believing: Some affective influences on belief. *Emotions and beliefs: How feelings influence thoughts*, pages 10–44.

Cohen, J. (1992). *An Essay on Belief and Acceptance*. Oxford University, Oxford.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4:317–370.

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

Coltheart, M., Langdon, R., and McKay, R. (2011). Delusional belief. *Annual review of psychology*, 62:271–298.

Covington, M. A., He, C., Brown, C., Naci, L., McClain, J. T., Fjordbak, B. S., Semple, J., Brown, J., et al. (2005). Schizophrenia and the structure of language: the linguist's view. *Schizophrenia Research*, 77(1):85–98.

Currie, G. (2000). Imagination, delusion and hallucination. In *Pathologies of Belief*, pages 168–183. Blackwell, Oxford.

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

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

Currie, G. and Ravenscroft, I. (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford University Press, New York.

Dancy, J. (2002). *Practical Reality*. Oxford University Press.

Davidson, D. (1982a). Psychology as philosophy. In *Essays on Actions and Events*, pages 229–238. Oxford University Press, Oxford.

Davidson, D. (1982b). Rational animals. *Dialectica*, 36:317–28.

Davidson, D. (1990). The structure and content of truth. *The Journal of Philosophy*, 87(6):279–328.

Davidson, D. (2001). How is weakness of the will possible? *Essays on actions and events*, 1(9):21–43.

Davies, M. and Coltheart, M. (2000). Introduction: pathologies of belief. *Mind & Language*, 15:1–46. 1.

Davies, M., Coltheart, M., Langdon, R., and Breen, N. (2001). Monothematic delusions: toward a two-factor account. *Philosophy, Psychiatry, & Psychology*, 8(2–3):133–158.

de Finetti, B. (1970). *Theory of Probability*. New York: John Wiley.

de Jong, P. and Muris, P. (2002). Spider phobia: Interaction of disgust and perceived likelihood of involuntary physical contact. *Journal of Anxiety Disorders*, 16(1):51–65.

de Jong, P., van den Hout, M., Rietbroek, H., and Huijding, J. (2003). Dissociations between implicit and explicit attitudes toward phobic stimuli. *Cognition and Emotion*, 17(4):521–545.

de Sousa, R. (1971). How to give a piece of your mind: or, the logic of belief and assent. *Philosophical Review*, XXXV:52–79.

Dennett, D. (1978a). *Brainstorms: Philosophical Essays on Mind and Psychology*. Bradford Books, Cambridge.

Dennett, D. (1978b). How to change your mind. In *Brainstorms: Philosophical Essays on Mind and Psychology*, chapter 16, pages 300–309. Bradford Books.

Dennett, D. (1987a). Brain writing and mind reading. In *Brainstorms: Philosophical Essays on Mind and Psychology*, pages 39–50. Harvester Press, Brighton.

Dennett, D. (1987b). *The Intentional Stance*. The MIT Press, Cambridge, MA.

Dennett, D. (1989). Mid-term examination: Compare and contrast. In *The Intentional Stance*. MIT Press.

Dennett, D. (2007). *Breaking the Spell: Religion as a Natural Phenomenon*. Penguin Books.

Dennett, D. C. (1969). *Content and Consciousness*. Routledge and Kegan Paul, London.

Dennett, D. C. (1991). Real patterns. *Journal of Philosophy*, 88(1):27–51.

Dennett, D. C. and LaScola, L. (2010). Preachers who are not believers. *Evolutionary Psychology*, 8(1):122–150.

Descartes, R. (1641). *Meditations on First Philosophy with Selections from the Objections and Replies. Trans. John Cottingham*. Cambridge University Press, Cambridge.

Douven, I. (2008). Underdetermination. In Psillos, S. and Curd, M., editors, *The Routledge companion to philosophy of science*, chapter 27, pages 292–302. Psychology Press.

Edwards, W. (1954). The theory of decision making. *Psychological bulletin*, 51(4):380.

Egan, A. (2008). Seeing and believing: Perception, belief formation and the divided mind. *Philosophical Studies*, 140(1):47–63.

Egan, A. (2009). Imagination, delusion, and self-deception. In Bayne, T. and Fernández, J., editors, *Delusion and self-deception: affective and motivational influences on belief formation*, pages 263–280. Psychology Press, New York.

Eilan, N. (2000). On understanding schizophrenia. In Zahavi, D., editor, *Exploring the Self*, volume 23 of *Advances in Consciousness Research*, pages 97–113. John Benjamins Publishing Company, Amsterdam.

Engel, P. (1998). Believing, holding true, and accepting. *Philosophical Explorations*, 1(2):140–151.

Engel, P. (2000). Introduction: the varieties of belief and acceptance. In Engel, P., editor, *Believing and Accepting*, number 83 in Philosophical Studies Series, pages 1–30. Kluwer Academic Publisher.

Eriksson, L. and Hájek, A. (2007). What are degrees of belief? *Studia Logica*, 86(2):185–215.

Evans, J. S. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology*, 87:223–40.

Evans, J. S. B. T. (2009). *Biases in human reasoning: causes and consequences*. Oxford University Press, Oxford.

First, M., editor (2000). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision*. American Psychiatric Association.

Flaum, M., Arndt, S., and Andreasen, N. C. (1991). The reliability of 'bizarre' delusions. *Comprehensive Psychiatry*, 32:59–65.

Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.

Fodor, J. A. and Lepore, E. (1992). *Holism: A Shopper's Guide*. Blackwell.

Frankish, K. (2004). *Mind and Supermind*. Cambridge University Press.

Frankish, K. (2009). Delusions: a two-level framework. In Broome, M. R. and Bortolotti, L., editors, *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, International Perspectives in Philosophy and Psychiatry, pages 269–285. Oxford University Press.

Frith, C. and Johnstone, E. (2003). *Schizophrenia: A Very Short Introduction*. Oxford University Press, Oxford.

Frith, C. D. (1992). *The cognitive psychology of schizophrenia*. The Psychology Press, Brighton.

Frith, C. D., Blakemore, S.-J., and Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355(1404):1771–1788.

Fuchs, T. (2005). Delusional mood and delusional perception–a phenomenological analysis. *Psychopathology*, 38(3):133–139.

Fuentenebro, F. and Berrios, G. E. (1995). The predelusional state: A conceptual history. *Comprehensive Psychiatry*, 36(4):251–259.

Fulford, K. W. M. (1989). *Moral Theory and Medical Practice*. Cambridge University Press, Cambridge.

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

Garety, P. A. and Hemsley, D. R. (1997). *Delusions: Investigations Into The Psychology Of Delusional Reasoning*. Maudsley Monographs. Psychology Press.

Gendler, T. (2008a). Alief in action (and reaction). *Mind and Language*, 23(5):552–585.

Gendler, T. S. (2008b). Alief and belief. *Journal of Philosophy*, 105(10):634–663.

Gilbert, D. (1991). How mental systems believe. *American psychologist*, 46(2):107.

Gilbert, D. T., Tafarodi, R. W., and Malone, P. S. (1993). You can't not believe everything you read. *Journal of personality and social psychology*, 65(2):221.

Goldman, A. (1989). Interpretation psychologized. *Mind and Language*, 4(3):161–85.

Goldman, A. (1993a). The psychology of folk psychology. *Behavioral and Brain Sciences*, 16:15–28.

Goldman, A. (2006). *Simulating Minds*. Oxford University Press, Oxford.

Goldman, A. I. (1993b). *Philosophical applications of cognitive science*, volume 153. Westview Press Boulder.

Grandy, R. (1973). Reference, meaning, and belief. *Journal of Philosophy*, 70(14):439–452.

Haack, S. (1996). Analyticity and logical truth in the roots of reference. In *Deviant Logic, Fuzzy Logic*, pages 214–228. The University of Chicago Press.

Harman, G. (1976). Practical reasoning. *Review of Metaphysics*, 29:431–63.

Hawthorne, J. (2009). Superficialism in ontology. In *Metametaphysics: New Essays on the Foundations of Ontology*, pages 213–230. Oxford University Press.

Heil, J. (1989). Minds divided. *Mind*, 98(392):571–583.

Heil, J. (1994). Going to pieces. In Graham, G. and Stephens, G. L., editors, *Philosophical Psychopathology*, pages 111–134. MIT Press.

Hirstein, W. (2005). *Brain Fiction: Self-Deception and the Riddle of Confabulation*. MIT Press.

Holton, R. (forthcoming). Intention as a model for belief. In *Rational and Social Agency: Essays on the Philosophy of Michael Bratman*. Oxford University Press.

Hume, D. (2000). *A Treatise of Human Nature*. Oxford University Press.

Ichikawa, J., Jarvis, B., and Rubin, K. (2012). Pragmatic encroachment and belief-desire psychology. *Analytic Philosophy*, 53(4):327–343.

Jaspers, K. (1913/1963). *General Psychopathology*. University of Chicago Press, Chicago.

Junginger, J., Barker, S., and Coe, D. (1992). Mood theme and bizarreness of delusions in schizophrenia and mood psychosis. *Journal of Abnormal Psychology*, 101:287–92.

Kahneman, D., Slovic, P., and Tversky, A., editors (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kaplan, M. (1998). *Decision theory as philosophy*. Cambridge University Press.

Kendler, K. S., Glazer, W. M., and Morgenstern, H. (1983). Dimensions of delusional experience. *American Journal of Psychiatry*, 140:466–9.

Kirk, S. A. and Kutchins, H. (1992). *The Selling of the DSM: The Rhetoric of Science in Psychiatry*. Transaction Publishers, New Brunswick.

Klee, R. (2004a). Delusional content and the public nature of meaning: Reply to the other contributors. *Philosophy, Psychiatry, and Psychology*, 11(1):95–99.

Klee, R. (2004b). Why some delusions are necessarily inexplicable beliefs. *Philosophy, Psychiatry, and Psychology*, 11(1):25–34.

Kriegel, U. (2012). Moral motivation, moral phenomenology, and the alief/belief distinction. *Australasian Journal of Philosophy*, 90(3):469–486.

Laing, R. D. (1969). *The Divided Self*. Pelican, London.

Landsburg, S. (2009). *The Big Questions: Tackling the Problems of Philosophy with Ideas from Mathematics, Economics, and Physics*. Free Press.

Langdon, R. and Coltheart, M. (2000). The cognitive neuropsychology of delusions. *Mind & Language*, 15(1):184–218.

Lewis, D. (1970). How to define theoretical terms. *Journal of Philosophy*, 67:427–46.

Lewis, D. (1974). Radical interpretation. *Synthese*, 27(July–August):331–344.

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

Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61(December):343–377.

Lewis, D. (1984). Putnam's paradox. *Australasian Journal of Philosophy*, 62(3):221–236.

Macintosh, J. (1995). Belief-in. In Honderich, T., editor, *The Oxford Handbook of Philosophy*. Oxford University Press.

Maher, B. A. (1974). Delusional thinking and perceptual disorder. *Journal of individual psychology*.

Maher, B. A. (1999). Anomalous experience in everyday life: its significance for psychopathology. *The Monist*, 82:547–570.

Maher, P. (1990). Acceptance without belief. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pages 381–392.

Mallon, R. and Stich, S. P. (2000). The odd couple: The compatibility of social construction and evolutionary psychology. *Philosophy Of Science*, 67(1):133–154.

Manley, D. (2009). Introduction: A guided tour of metametaphysics. In Chalmers, D., Manley, D., and Wasserman, R., editors, *Metametaphysics: New Essays in the Foundations of Ontology*, pages 1–37. Oxford University Press.

Marková, I. S. and Berrios, G. E. (1992). The meaning of insight in clinical psychiatry. *The British Journal of Psychiatry*, 160(6):850–860.

Mayer, B., Merckelbach, H., and Muris, P. (2000). Self-reported automaticity and irrationality in spider phobia. *87*, Psychological Reports(2):395–405.

McGinn, C. (2005). *Mindsight: Image, Dream, Meaning*. Harvard University Press, Cambridge.

McHugh, P. R. and Slavney, P. R. (1998). *The Perspectives of Psychiatry*. The Johns Hopkins University Press, second edition.

McLaughlin, B. P. (2006). Is role-functionalism committed to epiphenomenalism? *Journal of Consciousness Studies*, 13(1-2):1–2.

McLaughlin, B. P. (2009). Monothematic delusions and existential feelings. In Bayne, T. and Fernández, J., editors, *Delusion and Self-Deception*, pages 139–164. Psychology Press, New York.

Mele, A. R. (2001). *Self-Deception Unmasked*. Princeton University Press, Princeton, New Jersey.

Mercier, H. and Sperber, D. (2009). Intuitive and reflective inferences. In Evans, J. and Frankish, K., editors, *In Two Minds: Dual Processes and Beyond*. Oxford Univeristy Press.

Mishara, A. L. (2010). Klaus Conrad (1905–1961): Delusional mood psychosis, and beginning schizophrenia. *Schizophrenia Bulletin*, 36(1):9–13.

Morris, E. (2010). The anosognosic's dilemma: Something's wrong but you'll never know what it is (part 4). New York Times, Opinionator post of June 23.

Mullen, P. (1979). Phenomenology of disordered mental function. In Hill, P., Murray, R., and Thorley, G., editors, *Essentials of post-graduate psychiatry*, pages 25–54. Academic.

Murphy, D. (2005). *Psychiatry in the Scientific Image*. MIT Press.

Murphy, D. (2010a). Explanation in psychiatry. *Philosophy Compass*, 5(7):602–610.

Murphy, D. (2010b). Philosophy of psychiatry. *Stanford Encyclopedia of Philosophy*.

Murphy, D. (2012). The folk epistemology of delusions. *Neuroethics*, 5(1):19–22.

Murphy, D. (2013). Delusions, modernist epistemology and irrational belief. *Mind and Language*, 28(1):113–124.

Nagel, J. (2012). Gendler on alief. *Analysis*, 72(4):774–788.

Needham, R. (1972). *Belief, Language, and Experience*. Oxford,Blackwell.

Nichols, S. and Stich, S. P. (2000). A cognitive theory of pretense. *Cognition*, 74(2):115–147.

Nichols, S., Stich, S. P., Leslie, A. M., and Klein, D. B. (1996). Varieties of off-line simulation. In Carruthers, P. and Smith, P., editors, *Theories of Theories of Mind*, pages 39–74. Cambridge University Press.

Nisbett, R. and Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3):231–59.

Oltmanns, T. F. (1988). Approaches to the definition and study of delusions. In Oltmanns, T. F. and Maher, B. A., editors, *Delusional Beliefs*, pages 3–12. Wiley.

Papineau, D. (1991). *Reality and Representation*. Blackwell.

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

Quine, W. V. O. (1960). Carnap and logical truth. *Synthese*, 12(4):350–374.

Quine, W. V. O. and Ullian, J. S. (1970). *The Web of Belief*. Random House, New York.

Radden, J. (2010). *On Delusion*. Routledge.

Ramsey, F. P. (1931). Truth and probability. In Braithwaite, R. B., editor, *The Foundations of Mathematics and Other Logical Essays*, pages 156–98. Routledge and Kegan Paul, London.

Ratcliffe, M. (2005). The feeling of being. *Journal of Consciousness Studies*, 12(8–10):43–60.

Ratcliffe, M. (2007). *Rethinking Commonsense Psychology: a critique of folk psychology, theory of mind, and simulation*. New Directions in Philosophy and Cognitive Science. Palgrave Macmillan.

Ratcliffe, M. (2008). *Feelings of Being: phenomenology, psychiatry, and the sense of reality*. International Perspectives in Philosophy and Psychiatry. Oxford University Press.

Rawling, P. (2003). Radical interpretation. In Ludwig, K., editor, *Donald Davidson*, Contemporary Philosophers in Focus, pages 85–112. Cambridge University Press.

Reber, A. S. (1993). *Implicit learning and tacit knowledge*. Oxford Univeristy Press, Oxford.

Recanati, F. (2000). The simulation of belief. In Engel, P., editor, *Believing and Accepting*, number 83 in Philosophical Studies Series, pages 267–298. Kluwer Academic Publishers.

Rey, G. (1988). Toward a computational account of akrasia and self-deception. In McLaughlin, B. P. and Rorty, A. O., editors, *Perspectives on Self-Deception*, pages 264–296. University of California Press.

Rey, G. (2007). Meta-atheism: Religious avowal as self-deception. In Antony, L. M., editor, *Philosophers Without Gods:*, pages 243–265. Oxford University Press.

Rhodes, J. and Gipps, R. (2011). Delusions and the non-epistemic foundations of belief. *Philosophy, Psychiatry, & Psychology*, 18(1):89–97.

Rhodes, J. and Gipps, R. G. T. (2008). Delusions, certainty, and the background. *Philosophy, Psychiatry, and Psychology*, 15(4):295–310.

Rose, D., Buckwalter, W., and Turri, J. (manuscript). When words speak louder than actions: Delusion, belief, and the power of assertion.

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, and Psychology*, 11(1):71–80.

Schwarz, N. and Clore, G. L. (1996). Feelings and phenomenal experiences. *Social psychology: Handbook of basic principles*, 2:385–407.

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

Schwitzgebel, E. (2002). A phenomenal, dispositional account of belief. *Noûs*, 36:249–275.

Schwitzgebel, E. (2006). Belief. In *Stanford Encyclopedia of Philosophy*.

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

Scott, F. J., Baron-Cohen, S., and Leslie, A. (1999). if pigs could fly: A test of counterfactual reasoning and pretence in children with autism. *British Journal of Developmental Psychology*, 17(3):349–362.

Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press.

Sider, T. (2009). Ontological realism. In Chalmers, D., Manley, D., and Wasserman, R., editors, *Metametaphysics: New Essays in the Foundations of Ontology*, pages 384–423. Oxford University Press.

Silva, J. A., Leong, G. B., Weinstock, R., and Boyer, C. L. (1989). Capgras syndrome and dangerousness. *Journal of the American Academy of Psychiatry and the Law Online*, 17(1):5–14.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119(1):3–22.

Smith, E. R. and Collins, E. C. (2009). Dual-process models: A social psychological perspective. In Evans, J. S. B. T. and Frankish, K., editors, *In Two Minds: Dual Processes and Beyond*. Oxford University Press, Oxford.

Sober, E. (1978). Psychologism. *Journal for the Theory of Social Behavior*, 8:165–191.

Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach*. Blackwell, Oxford.

Sperber, D. (2000). Intuitive and reflective beliefs. In Engel, P., editor, *Believing and Accepting*, pages 243–266. Kluwer Academic Publishers.

Spitzer, M. (1990). On defining delusions. *Comprehensive Psychiatry*, 31:377–97.

Stalnaker, R. (1984). *Inquiry*. MIT Press, Cambridge, Mass.

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

Stich, S. (1983). *From Folk Psychology to Cognitive Science*. MIT Press, Cambridge, Mass.

Stich, S. (1985). Could man be an irrational animal? *Synthese*, 64(1):115–35.

Stich, S. (1988). Reflective equilibrium, analytic epistemology and the problem of cognitive diversity. *Synthese*, 74(3):391–413.

Stich, S. (1996). Deconstructing the mind. In *Deconstructing the Mind*, pages 3–90. Oxford University Press.

Stich, S. and Ravenscroft, I. (1994). What is folk psychology? *Cognition*, 50:447–68.

Thagard, P. and Nisbett, R. E. (1983). Rationality and charity. *Philosophy of Science*, 50(2):250–267.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In Evans, J. S. B. T. and Frankish, K., editors, *In Two Minds: Dual Processes and Beyond*, chapter 8, pages 171–196. Oxford University Press, Oxford.

Thornton, T. (2007). *Essential Philosophy of Psychiatry*. Oxford University Press.

Thornton, T. (2008). Why the idea of framework propositions cannot contribute to an understanding of delusions. *Phenomenology and the Cognitive Sciences*, 7(2):159–175.

Tumulty, M. (2011). Delusions and dispositionalism about belief. *Mind and Language*, 26(5):596–628.

Tumulty, M. (2012). Delusions and not-quite-beliefs. *Neuroethics*, 5(1):29–37.

Tuomela, R. (2000). Belief versus acceptance. *Philosophical Explorations*, 3(2):122–137.

van Fraassen, B. C. (1980). *The Scientific Image*. Oxford University Press, USA, Oxford.

Velleman, J. D. (1992). The guise of the good. *Noûs*, 26(1):3–26.

Weatherson, B. (2003). What good are counterexamples? *Philosophical Studies*, 115(1):1–31.

Weatherson, B. (2009). David lewis. In *Stanford Encyclopedia of Philosophy*.

Wedgwood, R. (2002). The aim of belief. *Philosophical Perspectives*, 16(s16):267–97.

Wessely, S., B. A. R. A. e. a. (1993). Acting on delusions. i: prevalence. *British Journal of Psychiatry*, 163:69–76.

Williams, B. (1973). Deciding to believe. In *Problems of the Self*, pages 136–51. Cambridge University Press.

Williams, J. R. G. (2007). Eligibility and inscrutability. *Philosophical Review*, 116(3):361–399.

Winters, K. C. and Neale, J. M. (1983). Delusions and delusional thinking in psychotics: a review of the literature. *Clinical Psychology Review*, 3:227–53.

Wittgenstein, L. (1975). *On Certainty*. Wiley-Blackwell.

Wright, C. (2004). Warrant for nothing (and foundations for free)? *Aristotelian Society Supplementary Volume*, 78(1):167–212.

Young, A. (1998). *Face and Mind*. Oxford University Press, Oxford.

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

Young, G. (2007). Clarifying "familiarity": Phenomenal experiences in prosopagnosia and the capgras delusion. *Philosophy, Psychiatry, & Psychology*, 14(1):29–56.

Zynda, L. (2000). Representation theorems and realism about degrees of belief. *Philosophy of Science*, 67:45–69.

Vita

Richard Dub

2013 Ph. D. in Philosophy, Rutgers University

2006 M. A. in Philosophy, Tufts University

2002 B. A. in English Literature, McGill University

2008-2011 Teaching assistant, Department of Philosophy, Rutgers University

2013 "Dennett's Rationality Assumption". Forthcoming in *Content and Consciousness Revisited*, eds. Muoz-Surez, C. and De Brigard, F., Springer.