Belief, Reasons, and Irrationality

Sophie Edwards

University College London

PhD

For Max

In gratitude

and promise

I, Sophie Edwards, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

Abstract

In this thesis, we explore the question: *What is a belief?* We do so by considering the folk psychological concept of belief and attempting to unearth some constitutive features of it. We argue that, according to this concept, there is a significant relationship between belief and reasons: one which reveals that *beliefs aim at truth*, as Bernard Williams (1973) once famously put it. We argue for a particular interpretation of this claim, according to which it is to be understood as follows:

(R): It is constitutive of belief that *if* it is consciously regulated, it is so-regulated solely for truth;

and

(C): It is constitutive of belief that it is correct if and only if its contents are true.

We maintain that (C) explains why it is that (R) is true. So, belief is at base a normative concept: the question as to why one holds a particular belief can always be raised. We then explore two irrational phenomena – self-deception and delusion – and further unravel what (R) and (C) involve, as well as shedding some light on the phenomena themselves. We argue for a position we call *doxastic minimalism* about self-deception, according to which, in the paradigm case, the self-deceiver holds neither their undesired belief that *p* nor their desired belief that ~*p*. This is because they do not have attitudes to these contents that meet conditions (R) and (C). Similarly, we argue that although cases of delusion vary significantly, in some extreme cases, the subject does not seem to relate to the content of their delusion in a manner that meets (R) and (C), and hence ought not to be attributed a belief in such contents.

Table of Contents

Acknowledgements

I would like to thank my supervisor Lucy O'Brien for her support and encouragement, but most of all for her substantial philosophical input and her ever-practical advice. From the philosophical community I would also like to thank Robert Audi, Matthew Boyle – whose work I find to be a constant inspiration – Fiona Leigh, Chris Peacocke, Eric Schwitzgebel; not to mention José Zalabardo for telling me about his pile of rejections and Conor McHugh for being the one not to reject me.

I would also like to thank the Arts and Humanities Research Council (AHRC) for providing full funding for this project.

For a much older debt I would like to thank my A-level philosophy teachers Mr Buxton and Ms Smith – wherever they are now – for enabling my first experiences of real intellectual excitement.

I would also like to thank my family: my mum for her constant love and support, my sister for her understanding, and my newly acquired in-laws (including Adam) for welcoming me so warmly into their family. In particular, I would like to thank John and Kathy Edwards for their unflinching support and for providing us with a place to live whilst I wrote-up this thesis. Thank you also to Naomi Weston for teaching me that true friendship reveals itself over many years.

Special thanks also go to my godparents Alan and Eileen Pike, and Meg Leitch for enriching my life with countless musical outings during the time this work was carried out. Also, thanks to my godfather for his constant presence, reassurance, and sage advice.

Finally, my love and thanks to my husband Max Edwards for showing me what it is and will be to lead a life together as philosophers.

Introduction

This thesis is driven, broadly speaking, by the desire to answer the following question: *What is a belief?* But this question has different senses, and consequently can be approached in different ways. At what is perhaps the most basic level, this question can be understood as a question about the *metaphysical* status of belief: are beliefs identical to certain brain states, or certain sets of behavioural dispositions, or are they perhaps merely convenient fictions for the meantime that according to a final neuroscience will turn out not really to exist at all? We will not approach the question by trying to answer it in this metaphysical sense – at least not directly. Rather, we will approach the question as to what a belief is in the sense in which it demands conceptual analysis of the folk psychological concept *belief*.

Folk psychology, by its very nature, is an informal set of ideas. It involves a collection of all of the kinds of things ordinary people say about belief – when they do say anything – as well as the propositions about belief that we can think of as entailed by ordinary people's willingness to attribute beliefs under certain conditions. Now, a conceptual analysis of the folk concept *belief* does not involve a mere cataloguing of everything any ordinary person has ever said or implied about it. For a start, it is part of the folk psychological concept of belief that the term *belief* can be misused. Rather, the project we undertake here involves a recognition of the *patterns* in the kinds of things people say about belief, and what most people think of as non-negotiable truths about the way in which the concept is to be used. Working with these patterns and non-negotiables, we will then attempt to extract some significant features of the way in which the concept *ought* to be used by, and according to, us, the folk. So, the result will be the description of a normative claim implicit in folk psychology. Now, because of this fact, and because the information about what people say and imply about belief we begin with is such a rag-bag set, inevitably some of these things will not be accommodated by our result. But this in itself will not disprove it. Ultimately, our conception of belief will stand or fall on the grounds of its being the best possible systematisation of the folk data, that is, whether it – in the simplest and most coherent manner possible – does more justice to the most important of our platitudes about belief than any other theory.

Now, this project makes sense independently of a complete theory of what beliefs are over and above their role in folk psychology. We can think about the rules for the use of the concept *belief* in the folk psychological language game, independently of asking whether this concept actually refers to anything outside of this talk, for example. But, of course, what one says about how the folk think their concept ought to be used will impact upon the question as to its metaphysical nature, since what the concept *is* will impact upon the question as to whether or not it refers. In fact, we cannot look for beliefs in the world if we do not have *some* idea of what it is we are looking for. Now, when we go looking, it will be an open question as to how much of our best systematisation of our folk concept need turn out to refer for us to say that we have discovered *beliefs*. Perhaps if what we find is just slightly different from what we were looking for, then we ought to revise our folk concept accordingly. However, if what we find in the natural world differs too greatly from what we set out to look for, then referring to what we have found as a *belief* seems suspect. Rather, it may be better to conclude that there are no beliefs. With this distinction between the metaphysical and the folk psychological approach to the question in mind, it should be clear that from now on when we talk about the nature of belief, we are talking about how things are according to the folk concept *belief*.

In order to find our way in to our conceptual analysis of belief, we will begin in chapter one by considering some cases in which it is not obvious what to say the person in question believes. In general, when thinking about a concept, this seems a sensible way to proceed, as it is precisely these kinds of cases that tend to reveal what is at issue between different analyses of the concept. We will consider two contemporary approaches to belief attribution. The first is due to Eric Schwitzgebel (e.g. 2001b), who claims that it is often appropriate to say that people are *in-between* believing. The second is propounded by Tamar Gendler (e.g. 2008b), who denies that people are ever in-between believing in this way: rather, she claims that we often possess what she calls *belief-discordant aliefs*. We will see that what the debate between Schwitzgebel and Gendler amounts to is the age-old debate concerning the significance of the relationship between belief and reasons, with Gendler affirming the significance of this relationship, and Schwitzgebel denying it.[1] By the end of the

---

[1] David Owens (2000, pp. 1-2) offers a helpful characterisation of this debate, on which one pole is typified by Immanuel Kant, René Descartes, and John Locke, which conceives of belief as subject to reason in some significant sense, and one typified by David Hume (1978, p. 183), who thinks that

first chapter, we will have established that although the details of Gendler's account render it unconvincing, her claim that there is a significant relationship between belief and reasons seems to be an important one. The rest of the thesis should be understood as an explication and defence of this claim.

In chapter two, we will begin our exploration of the relationship between belief and reasons by considering two apparent features of belief. The first is what Nishi Shah (2003, 2005) has called the *transparency* of doxastic deliberation. Transparency is the claim that when explicitly considering *whether to believe that p* the phenomenology is as of this question immediately (that is, non-inferentially) giving way to the question *whether p*. This phenomenon seems to support our sense that the relationship between belief and reasons is a significant one. The second, which we will refer to as *heterogeneity*, is the truism that regardless of what one might think about one's beliefs, they are often influenced – outside of conscious awareness – by factors other than reasons in support of them, such as what one desires to believe, for example. Heterogeneity seems to put pressure on our sense that reasons and belief are importantly interrelated. However, we will argue that Bernard Williams' (1973) famous claim that *beliefs aim at truth*[2] can accommodate both of these ideas about belief. According to our interpretation of this claim, it is constitutive of belief that – regardless of influences on it outside of consciousness – *if* it is consciously regulated (formed, upheld, revised, and extinguished), it is so-regulated exclusively for truth. This is what we will refer to as the *regulation condition*. Furthermore, what explains this regulation condition is the fact that belief is a fundamentally normative notion: it is constitutive of belief that it is *correct* if and only if its contents are true. Given that this is the case, we can explain transparency as follows. The question *whether to believe that p* amounts to the question *should I believe that p*. So, normativity is implicit in the very deliberative standpoint itself. Given that this is the case, and given that in deliberating about whether one should believe that *p* one understands that it is constitutive of belief that it is correct if and only if its contents are true, one understands that the only relevant considerations are those pertaining to the truth of *p*. Having thus explained transparency and established that the relationship between

_____

"belief is more properly an act of the sensitive, than of the cogitative part of our natures" and that the "experimental reasoning" which governs belief "is nothing but a species of instinct or mechanical power, that acts in us unknown to ourselves" (Hume, 1975, p. 108).
[2] We shall be assuming a correspondence theory of truth throughout.

belief and reasons reveals that beliefs aim at truth in both the regulation and the correctness senses, we will then proceed to examine each of these conditions in further detail.

In chapter three, we will explore the regulation condition. We will begin by drawing attention to two further features of the manner in which beliefs are consciously regulated. The first of these is what we will call *exclusivity*. Exclusivity is the claim that when one considers *whether to believe that p*, one can only think of oneself as moved to form the belief that *p* on the basis of epistemic reasons. The second, which we will call *demandingness*, is the claim that when one considers *whether to believe that p*, one can only think of oneself as moved to form the belief that *p* on the basis of *sufficient* epistemic reason that *p*. We will explore Owens' (2003) argument that we cannot conceive of the regulation condition in terms of the doxastic deliberator possessing a literal aim or intention, because deliberation involving ordinary, non-epistemic aims does not exhibit exclusivity like doxastic deliberation. We will explore an alternative model for the doxastic deliberator's aim: they aim to believe that *p* if and only if they are *satisfied* that *p*. We maintain that non-epistemic aims of this form, such as *to try the restaurant if and only if one is satisfied that it received good reviews* do prompt deliberation that exhibits a non-epistemic analogue of exclusivity – one can only consider whether one is satisfied in such a case. However, we will conclude that the doxastic deliberator cannot possess the literal aim *to believe that p if and only if they are satisfied that p*, because in the epistemic case, such an aim would be circular. Being consciously satisfied that *p* simply is to consciously believe that *p*. So, we will conclude that the regulation condition, and with it, both exclusivity and demandingness, is to be understood in terms of belief itself – not the believer – as aiming at truth in the sense that coming to consciously believe that *p is* coming to be consciously satisfied that *p*.

In chapter four, we will then consider the correctness condition. We will argue that we can derive two subsequent epistemic norms from the correctness condition, which instruct us as to when a belief is rational: *normative exclusivity* and *normative demandingness*. These combine to yield:

(Ra): One should believe that *p* if and only if one does so solely on the basis of sufficient objective epistemic reason in favour of *p*.

We will consider two main objections to the idea that both the correctness condition and, derivatively, (Ra) are constitutive of belief. The first of these is the *pragmatist*'s claim that beliefs are answerable to practical norms. We will argue that the intuitions pragmatists attempt to provoke can be explained by distinguishing between when *it is rational to believe that p* and *when it is rational to try to get oneself to believe that p*. The former, unlike the latter, is answerable solely to epistemic norms. We will then consider the contention that the correctness condition and its derivative norms cannot be understood to be constitutive of belief because we do not have the right kind of control over our beliefs to render the manner in which we are held responsible for them coherent. We will argue that we can make sense of the manner in which we are held responsible to epistemic norms if we conceive of the kind of control we have over our beliefs in terms of our capacity to *reflect* upon our reasons for our beliefs. At any given time, we are free to reflect upon why we hold a belief and regulate it accordingly: this is why, at any given time, we are to be held responsible for believing as we do.

In chapters five and six we will then turn to two phenomena that – like the problem cases we began with – put pressure on the limits of our concept of belief. We will examine what the account of belief we have been developing can teach us about these phenomena, and *vice versa*. We will begin with self-deception in chapter five. It is often thought that part of what it is to be self-deceived is to hold contradictory beliefs – he both unconsciously believes that his wife *is* having an affair, and consciously believes that it is *not* the case that she is having an affair. Against this, we will argue for a position that we will call *doxastic minimalism* about self-deception. This is the claim that, in the paradigm case, the self-deceiver holds neither their undesired belief that *p*, nor their desired belief that ~*p*. We will begin by presenting a negative case for this claim. We will examine various considerations – pertaining to the self-deceiver's behaviour, and the phenomenology of self-deception, for example – and we will argue that none of them demonstrate an explanatory need to attribute the self-deceiver either their undesired or their desired belief, considered independently from one another. Folk psychology is sufficiently rich to afford us many alternative concepts – such as *suspicion, anxiety, hope* etc. – that we can employ instead of belief. We will then consider the idea that we need attribute them *at least one* of these beliefs – perhaps to distinguish self-deception

from simple *ambivalence*, for example. However, we will maintain that, again, no explanatory need to attribute the self-deceiver at least one of the beliefs in question has been demonstrated. We will then proceed to make a positive case for doxastic minimalism by utilising the conception of belief we have been developing, and in so doing, we will learn more about this concept. We will argue that the self-deceiver cannot be said to hold their undesired belief, as they would have to be unable to reflect upon it during the time of their self-deception, and this is incompatible with our claim that we have reflective control over our beliefs. So, by examining self-deception we will learn that our account of belief precludes beliefs from being unconscious in the sense that we cannot reflect upon them. And, secondly, with respect to the self-deceiver's desired belief, we will argue that when they consciously consider their desired state of affairs, they cannot be understood to be regulating their thinking solely in accordance with truth as the regulation condition demands, because the phenomenology of self-deception precludes this conclusion. In the paradigm case, the self-deceiver feels unease surrounding the topic of their self-deception: this is incompatible with their coming to be satisfied that ~*p*, as the regulation condition demands.

Finally, in chapter six, we will turn to the phenomenon of delusion. Unlike in our treatment of self-deception, we will not be concerned to isolate paradigm cases of delusion, proceeding instead by enumerating, with reference to a number of cases, some of the features that most cases of delusion will exhibit to some extent. As with self-deception, our interest in delusion is an interest in belief attribution: we will be concerned to determine whether or not the deluded subject can be said to *believe* the content about which they are deluded. Whilst we will acknowledge that a number of central cases are plausibly regarded as species of irrational beliefs, we will argue that this *doxasticist* construal of delusion does not get a grip on certain more extreme cases. In such cases, we will suggest that the subject is better seen as periodically alienated from their nature as a rational agent and their status as a believer, rather than simply forming an irrational belief. Three considerations motivate this verdict: first, subjects in such cases show little or no tendency to rationally integrate the delusional content with their other beliefs, secondly, it puts a strain on our concept of a reason to describe them as taking themselves to have sufficient reason to rationally uphold the delusional content, and finally, we do not extend to such subjects the kind

of rational agency that is required for us to hold them rationally responsible, and hence for the correctness condition to hold. Thus, we find in deception a clarification and reaffirmation of the claim that belief is an essentially normative concept.

1. Juliet and the Precipice

Belief attribution is not always easy. In recent philosophical writing, there has been a flurry of discussion of certain cases in which it is in fact extremely contentious what the person in question believes. In this chapter, we will examine two such cases. The first is dubbed the *Implicit Racist* by Schwitzgebel (2010). The implicit racist – Juliet – is a Caucasian-American philosophy professor who "has critically examined the literature on racial differences in intelligence, and… finds the case for racial equality compelling. She is prepared to argue coherently, sincerely, and vehemently for equality of intelligence and has argued the point repeatedly in the past. And yet…[she] is systematically racist in most of her spontaneous reactions, her unguarded behaviour and her judgments about particular cases" (Schwitzgebel, 2010, p. 532). The second is a case that we shall call *Precipice*, versions of which were discussed by the early moderns (see e.g. David Hume (1978, p. 100), Michel de Montaigne (1957, p. 250)). Here is Gendler's (2008b, p. 634) contemporary example:

"In March 2007, 4000 feet above the floor of the Grand Canyon, a horseshoe-shaped cantilevered glass walkway was opened to the public. Extending 70 feet from the Canyon's rim, the Grand Canyon Skywalk soon drew hundreds of visitors each day…

'[who know]… what sensation is being promised…The promise is the dizzying thrill of vertigo.

And indeed, last week some visitors to this steel supported walkway anchored in rock felt precisely that. One woman, her left hand desperately grasping the 60-inch-high glass sides and the other clutching the arm of a patient security guard, didn't dare move toward the transparent center of the walkway.' (Rothstein, The New York Times, 19th May 2007)"

What are we to say that Juliet and the woman in Precipice *believe*? Does Juliet believe that the races are intellectually equal? Does the woman in Precipice believe that she is safe? In this chapter, we will see that what is really at issue when it comes

to adjudicating belief attribution in these difficult cases is the nature of the long-disputed relationship between belief and reasons.[3]

We will begin by outlining the four main options that one could take in response to such problem cases. We will focus on versions of two of these options, which are representative of the two poles of the debate over the relationship between belief and reasons. The first, adopted by Schwitzgebel (2001b, 2010), denies that there is any significant relationship between belief and reasons, and the second, adopted by Gendler, affirms the significance of this relationship.

Schwitzgebel's response is invited by *dispositionalism* about belief. Given that dispositionalists claim that all there is to believing that *p* is possessing a certain set of dispositions, if – as in our two cases – someone possesses only *some* of the requisite dispositions for a belief that *p*, the dispositionalist can simply maintain that the person in question is *in-between* believing that *p*.[4]

We will begin to cast doubt upon such a position by drawing attention to the fact that it is unable to explain the *pattern* discernible in the behaviour of the person in each case: all of their avowals and considered behaviour hang together, as does all of their unthinking behaviour. We will initially suggest that a folk-psychological account according to which the person believes that *p*, but has habitual and/or instinctual reactions otherwise can provide such an explanation. Their considered behaviour all hangs together as it is representative of what they believe, whereas their unthinking behaviour manifests some other folk psychological state.

We will then compare this kind of folk psychological approach to Gendler's (2008b, 2008c) recent proposal. According to Gendler, both Juliet and the woman in Precipice hold the relevant belief; they simply *alieve* otherwise. In order to investigate whether this proposal represents an improvement on our folk psychological one, we will explore the concepts of habit and instinct. We will see that, although our folk psychological concept of habit is a fairly robust one, our concept of an instinct is shrouded in controversy, because it is unclear whether it will

---

[3] For now, we will not explicitly define the concept of a reason, working instead with an intuitive sense of what it is for something to be a reason in favour of *p*'s truth. We will draw some important distinctions concerning types of reason at the beginning of the next chapter.

[4] We will talk throughout about believing that '*p*'. However, this should not be taken to imply a commitment to the idea that the contents of beliefs are propositions. '*p*' should simply be understood to stand for whatever the contents of belief turn out to be.

turn out to be helpful to talk about certain characteristics as *innate*. Furthermore, even if the concept of innateness does turn out to be a helpful one, whether the kind of behaviour evinced in Precipice results from an innate propensity is of course an empirical question that we cannot settle. However, we will argue that this does not affect the stability of the idea that a folk psychological explanation can cope with cases such as Precipice, as it is the concept of *fear* that is really doing the work in such a case, regardless of whether the propensity to such fear is acquired or innate. Having established our candidate explanations for both Implicit Racist and Precipice, we will argue that describing these cases in terms of alief affords no explanatory benefit.

Despite our disagreement with Gendler over the utility of the concept alief, we will see that we are nonetheless in agreement with her that the relationship between belief and reasons is a significant one. We will introduce the idea that it is *because* there is a significant relationship between belief and reasons that belief explains Juliet and the woman in Precipice's considered behaviour and not their unthinking behaviour. Their considered behaviour is reflective of what they understand their reasons to be, and hence of their beliefs, whereas their unthinking behaviour is not reflective of what they understand their reasons to be, and hence not of their beliefs. Rather, such behaviour is reflective of habit in Juliet's case, and fear in the case of Precipice. The rest of this thesis can then be understood as a defence and articulation of the idea that there is this kind of relationship between belief and reasons.

1.1 The Options

Assuming for the moment that believing is not a matter of degree, there appear to be four main options when it comes to characterising the beliefs of both Juliet and the woman in Precipice, where p is the belief that the races are intellectually equal for Juliet, and that she is safe for the woman in Precipice. In each case we might say either that:

(1)   She both believes that p and she believes that $\sim p$.

(2)   She believes that p and she does not believe that $\sim p$.

(3)   She believes that $\sim p$ and she does not believe that p.

(4)   She neither believes that p nor believes that $\sim p$.

In what follows, we will not explicitly discuss options (1) or (3), but our arguments in favour of (2) will of course represent arguments against accepting any of the three remaining options. We will explicitly pitch our arguments against Schwitzgebel's approach, which is a version of (4).

1.2 Schwitzgebel and In-Between Believing

Schwitzgebel (2010, p. 531) claims that in cases such as Implicit Racist and Precipice, there is no determinate fact about what the person in question believes. Rather:

"Such cases should be regarded as 'in-between' cases of believing, in which it's neither quite right to ascribe the belief in question nor quite right to say that the person lacks the belief."

Schwitzgebel's in-between believing approach is invited by his *dispositionalism* about belief. Roughly speaking, dispositionalists about belief claim that what it is to believe that *p is* to be disposed to behave as though *p* is the case. Following Gilbert Ryle (2000, p. 114), Schwitzgebel thinks that beliefs:

"…signify abilities, tendencies or pronesses to do, not things of one unique kind, but things of lots of different kinds."

In order to articulate this idea, he employs the notion of a *dispositional stereotype*, conceiving of a stereotype as:

"….a cluster of properties we are apt to associate with a thing, a class of things, or a property."

He does not wade into the complex debate as to the precise nature of a disposition, preferring instead simply to think of dispositions as:

"…characterized by means of conditional statements of the form: If condition C holds, then object O will (or is likely to) enter (or remain in) state S. O's entering S we may call the *manifestation* of the disposition, C we may call *condition of manifestation* of the disposition, and the event of C's obtaining we may call the *trigger*." (Schwitzgebel, 2002, p. 250)

Schwitzgebel does not enter into the debate as to the precise nature of a disposition as he does not need to in order to discuss his account of belief. For the same reason, we do not need to enter into this debate: we need simply grant Schwitzgebel that there are dispositions.

Schwitzgebel (2002, p. 252) does emphasise, however, that the dispositional stereotype associated with each belief includes not only what he calls *behavioural* dispositions, but also *phenomenal*, and *cognitive* dispositions. That is, as well as being disposed to behave in both verbal and non-verbal ways, he thinks that believing that *p* also disposes one to have certain phenomenological experiences, as well as to enter into other mental states that are not characterised entirely in terms of their phenomenology, such as drawing a further conclusion from one's belief.

Furthermore, a crucial assumption made by Schwitzgebel is what we will call his *egalitarianism*. Schwitzgebel assumes that all of the dispositions in the stereotype are *on a par* with one another: the believer's possession of no one disposition in the stereotype is of any greater significance than another when it comes to determining whether to attribute them the belief in question.

Given such a picture of belief, Schwitzgebel claims that if someone possesses *all* or a great deal of the dispositions in the stereotype associated with a certain belief, then it is appropriate to describe them as holding that belief; if none or very few, then it is appropriate to say that they do not hold the relevant belief. However, if they possess a fair few of the relevant abilities and tendencies but not others, we should say that they are in-between believing that *p*.

It is important to distinguish Schwitzgebel's position from one with a name that suggests affinity: Henry Price's (1969) claim that there are what he calls *half-beliefs*. Half-believing as Price understands it is importantly different from being in-between believing, in that a half-belief is:

"…something which is 'thrown off' when circumstances alter. In some sorts of contexts one is in a belief-like state with regard to a proposition, but in others one disbelieves it or just disregards it." (Price, 1969, p. 312)

In contrast, according to Schwitzgebel, in-between belief is not context relative. When one is in-between believing that *p*, one simply possesses some of the dispositions relevant to believing that *p* across all contexts.

Furthermore, Schwitzgebel's position must also be distinguished from one involving *degrees of belief*, or *credences*. Credences are famously associated with Frank Ramsey (1926), who claims that beliefs come in degrees between 0 and 1, also maintaining that they can be measured by betting behaviour.[5] So, for example, one might assign *p* a credence of .6, which Schwitzgebel would distinguish from a case in which one was in-between believing that *p* in terms of the dispositional set associated with each state. In the former case, one would have *all* or most of the dispositions associated with possessing a credence of .6 that *p*. For example, one would be disposed to say *I think that there is a .6 chance of p*, or perhaps, more colloquially, something along the lines of *I think that p is somewhat more likely than not*. In the case of being in-between believing that *p* and lacking this belief, however, one would have *some* of the relevant dispositions associated with believing that *p* and lack some of the others. So, one would either be disposed to say that *p* or refrain from so doing, for example, but one would not say *maybe p*. It is for this reason that Schwitzgebel rejects Darrell Rowbottom's (2007) suggestion that cases such as that of Juliet and Precipice can be conceived as involving degrees of belief. Rowbottom (2007, p. 134) insists that:

"…the changes [in Juliet's behaviour] can easily be accounted for by degrees of belief provided that the background information of… [Juliet] differ[s] in those situations…Change the context, and the (relevant conditional) degree of belief in *p* will change."

However, Schwitzgebel (2001b, p. 79) counters that "it is not a matter of degrees of belief fluctuating over time", such that:

"Juliet believes that all the races are intellectually equal when engaged in public debate, much less so or not at all when surveying her students on the first day of class." (Schwitzgebel, 2010, p. 543)

---

[5] It seems implausible that one's betting behaviour will line up perfectly with the probability one assigns to the truth of a particular proposition, because betting behaviour seems to reveal other attitudes as well – most obviously, one's attitude to risk.

Indeed, he claims that:

"…it seems possible for Juliet in a single moment both to be having a racist reaction and to be sincerely judging that the races are intellectually equal – for example, when she's having a racist reaction and trying to suppress it or when she's grading a black student's essay on intellectual equality, undervaluing the essay but regarding its conclusions as true. This is a possibility the shifting model gives us no means to accommodate." (Schwitzgebel, 2010, p. 543)

Rowbottom could perhaps insist that when a single act of Juliet's seems to reveal both a belief in the equality of the races and a belief to the contrary (or a lack of the relevant belief), Juliet simply believes .5 that the races are intellectually equal, for example. However, it seems as though Schwitzgebel (2001b, p. 79) is correct that Juliet does not:

"…have the kind of simple uncertainty that can be characterized in Bayesian terms as a unitary degree of belief – the kind of uncertainty one might have about tomorrow's stock prices or the value of an unseen card."

If Juliet simply has .5 confidence in the intellectual equality of the races in such a situation, then why isn't she simply disposed (*ceteris paribus*) to admit that she is not convinced one way or the other whether the races are intellectually equal?

So, in sum, Schwitzgebel claims that in cases such as Implicit Racist and Precipice, the person involved is in-between believing that *p* and lacking this belief, in so far as they possess a significant number of the behavioural dispositions relevant to the belief that *p*, whilst lacking others.

1.3 Tracing Patterns

However, as Schwitzgebel (2010, p. 534) says:

"Dispositional claims, like most generalizations, hold only *ceteris paribus*, 'all else being equal', or only against a defeasible set of background assumptions."

Let us consider a case that Rowbottom (2007) borrows from Ryle (1949, p. 129). Ryle claims that to believe that some ice is dangerously thin is to be disposed:

"[A] to be unhesitant in telling oneself and [B] others that it is thin, [C] in acquiescing in other people's assertions to that effect, [D] in objecting to statements to the contrary, [E] in drawing consequences from the original proposition, and so forth. But it is also [F] to be prone to skate warily, [G] to shudder, [H] to dwell in imagination on possible disasters and [I] to warn other skaters." (with Rowbottom's inserted letters)

Now, Rowbottom asks us to imagine an individual who has a certain subset of these dispositions. He has the disposition [A] to be unhesitant in telling himself that the ice is thin, and [E] in drawing consequences from the original proposition, [F] to be prone to skate warily, [G] to shudder, and [H] to dwell in imagination on possible disasters, and yet does not have the disposition [B] to tell others that the ice is dangerously thin, [C] to acquiesce in other people's assertions to that effect, [D] to object to statements to the contrary, or [I] to warn other skaters. Should we say that such a person is in-between believing that the ice is dangerously thin and lacking this belief? As Rowbottom (2007, p. 135) says:

"A credible explanation is that he is a nasty prankster who is sure that the ice is dangerously thin, but would delight in seeing someone else fall through it!"

Schwitzgebel (2010, p. 534) admits that if the person in question:

"…doesn't warn the other skaters out of *schadenfreude* or because he's blinded by the sun, that deviation from the typical dispositional manifestation counts not at all against ascribing him the belief that the ice is thin."

He doesn't warn the other skaters because *all else is not equal* – there is a certain "excusing condition" in place: he wants someone else to fall through.

Now, Schwitzgebel insists that there must be limits placed upon what can count as an excusing condition. He says:

"I'd say roughly, that when a candidate excusing condition would undermine the potential usefulness of the generalization, we should reject it as an excuser." (Schwitzgebel, 2010, p. 534)

Nonetheless, it seems that just as *schadenfreude* and being blinded by the sun can count as excusing conditions in the case of the ice skater, so there may be excusing

conditions in the case of Implicit Racist and Precipice, which account for why, although they hold the relevant belief, the person in question fails to exhibit some of the behavioural dispositions associated with that belief. In Juliet's case, it seems that a natural folk psychological explanation of her behaviour is that she believes that the races are intellectually equal, but she has recalcitrant *habits of reacting* otherwise (perhaps to be explained by her 1950s upbringing in the USA, for example). Similarly, there seems to be a ready folk psychological explanation of Precipice: the woman believes that she is safe – she would not go near the walkway otherwise – but she is *instinctually afraid* at such a great height. Neither the attribution of a habit in Juliet's case, nor an instinctual fear in Precipice undermine the usefulness of the generalisation pertaining to the relevant belief in each case.

So, how do we choose between Schwitzgebel's in-between believing interpretation of each case and what we have called the folk psychological explanation? One reason to choose the folk psychological explanation in each case is that it provides a *principled reason* as to why the person possesses exactly the dispositions that they do.[6] Let us consider Ryle's thin ice example again to make this point clear. Postulating that the skater believes that the ice is thin and yet is in the grip of the excusing condition that he wants someone to fall in explains why he has dispositions A, E, F, G, & H, and not B, C, D, & I. If the skater were in-between believing – which, remember, Schwitzgebel doesn't say he is – he could have *any* combination of the dispositions and lack any other combination; all that matters is that he has some and not others. However, Schwitzgebel resists claiming that the skater is in-between believing precisely because he doesn't just have a random combination and lack the others – there is a *pattern* to be traced in which dispositions he has and which he lacks, which is to be explained by the fact that he believes that the ice is thin, yet wants someone to fall in.

Surely, however, we should say something similar about Juliet and the woman in Precipice. In Juliet's case, for example, there is a *pattern* in her behaviour, which Schwitzgebel's hypothesis that she is in-between believing cannot capture. All of the things that she says upon reflection hang together with all of the things that she does in her considered moments, and all of her unconsidered reactions hang together. It is

---

[6] Thanks to Lucy O'Brien for initially suggesting this line of criticism.

not as if she has a seemingly random combination of the relevant dispositions. Now, unlike Schwitzgebel's position, our folk psychological account offers an explanation of this fact. Our account traces the precise structure of her dispositional profile by attributing to her the belief that the races are equal but the recalcitrant habit of reacting as if they are not. Similarly, in Precipice, the woman's considered behaviour – forcing herself to step out on to the walkway as well as to keep her eyes open, for example – all points in one direction, whereas her unconsidered behaviour all in another. This pattern can be captured by the hypothesis that she believes herself to be safe on the walkway, but is instinctually fearful of being at such a great height on a transparent surface.[7]

Furthermore, just as such folk psychological accounts can explain why Juliet and the woman in Precipice exhibit precisely the kinds of behaviour they do, folk psychology can of course also *predict* such patterns of behaviour for people in these kinds of circumstances. Schwitzgebel, however, would be left with no capacity at all to predict how someone would behave under such circumstances before the event.

1.4 Gendler and Alief

By way of further assessment of these proposed folk psychological explanations of both Implicit Racist and Precipice we will now turn to Gendler's alternative. Gendler (2008b, 2008c), *pace* Schwitzgebel, agrees that Juliet does believe that the races are intellectually equal, but, in addition, Gendler claims that she has a "belief-discordant alief" that points in the opposite direction. Similarly, the woman in Precipice does believe that she is safe according to Gendler, but she also has a belief-discordant alief that this is not the case. According to Gendler (2008b, p. 641), aliefs are:

"…*a*ssociative, *a*utomatic, and *a*rational. As a class, aliefs are states that we share with nonhuman *a*nimals; they are developmentally and conceptually *a*ntecedent to other cognitive attitudes that the creature may go on to develop. And they are typically also *a*ffect-laden and *a*ction generating."

Gendler (2008b, p. 642) defines paradigmatic alief as:

---

[7] It should be noted that we need not deny that there could be cases similar to Precipice in which the subject's fear is capable of causing their beliefs to change by prompting a re-evaluation on their part of their evidence. The present claim is simply that a *prima facie* plausible explanation of Precipice as Gendler presents it is that the woman's beliefs remain unchanged, but her instinctual fear is powerful enough to prevent those beliefs from playing their characteristic action-guiding roles.

"…a mental state with associatively linked content that is representational, affective and behavioural, and that is activated – consciously or nonconsciously – by features of the subject's internal or ambient environment. Aliefs may be either occurrent or dispositional."

So, according to Gendler, paradigmatic aliefs are (1) representational – they involve:

"…the representation of some object or concept or situation or circumstance, perhaps propositionally, perhaps nonpropositionally, perhaps conceptually, perhaps nonconceptually";

They are (2) affective – they involve "the experience of some affective or emotional state" – and they are (3) behavioural – they involve "the readying of some motor routine". Despite taking pains to spell out this four-place relation in this way, Gendler (2008b, p. 645) admits that she continues:

"…to waver on whether it would be better to think of… [alief] as two-place (*S* alieves *R*[epresentation]) rather than four-place (*S* alieves *R*[epresntation]-*A*[ffect]-*B*[ehaviour])."

Nonetheless, for brevity's sake, she often simply talks in terms of *S* alieving that *p*. So, one might say that in Implicit Racist, Juliet believes that the races are intellectually equal, yet alieves that they are not. The more extended version would be to say something along the following lines: Juliet believes that the races are intellectually equal, yet alieves "Other! Unequal! Avoid!" In Precipice, the extended version would be something like the woman believing that she was safe, whilst alieving "Really high up, long, long way down. Not a safe place to be! Get off"" (Gendler, 2008b, p. 635). From now on, we will simply talk about subjects alieving that *p*, because, first, it is unclear exactly what it is for a mental state to have a four-part content of the kind Gendler suggests, and, secondly, it seems that talking about subjects alieving that *p* captures the core of Gendler's position.

1.5 Alief and Folk Psychology

Now, Gendler does not propose to *replace* existing folk psychological concepts such as habit and instinct with her novel category; rather, she intends the concept of alief to *supplement* our existing vocabulary. She claims that whatever differences these

concepts might exhibit do not debar us from grouping them together under the rubric of alief. And, indeed, doing so:

"…brings out the connection between a number of otherwise apparently discrepant issues, and renders unmysterious a number of otherwise perplexing phenomena." (Gendler, 2008c, p. 553)

Before considering Gendler's arguments to this effect, let us first consider the nature of both habit and instinct, and their applicability to the cases we have considered. We will begin with habit. Ryle (2000, p. 42) outlines the contrast between what he calls "intelligent capacities" and habits as follows:

"When we describe someone as doing something by pure or blind habit, we mean that he does it automatically and without having to mind what he is doing. He does not exercise care, vigilance, or criticism. After the toddling-age we walk on pavements without minding our steps. But a mountaineer walking over ice-covered rocks in a high wind in the dark does not move his limbs by blind habit; he thinks what he is doing, he is ready for emergencies, he economizes in effort, he makes tests and experiments; in short he walks with some degree of skill and judgement… It is of the essence of merely habitual practices that one performance is a replica of its predecessors. It is of the essence of intelligent practices that one performance is modified by its predecessors… We build up habits by drill, but we build up intelligent capacities by training."

There are three features of habit that Ryle is drawing attention to here. One is that they prompt "automatic" behaviour – as he later says, one has been drilled in certain habits when one "can 'do them in his sleep', as it is revealing put" (Ryle, 2000, p. 42). As William James (1983, p. 119-121) says, "habit diminishes the conscious attention with which our acts are performed", such that "the attention may be wholly absorbed elsewhere". The second feature, related to this first, is that habits evince the same (or relevantly similar) behaviour every time such behaviour is enacted. The third is a fact about the manner in which habits are obtained: they are built up *via* drill or repetition.

A further fact about habit worthy of note is that it is not only human animals that have habits: it is also appropriate to describe the behaviour of many non-human animals as habitual, as James (1983, p. 125) claims:

"Riderless cavalry-horses, at many a battle, have been seen to come together and go through their customary evolutions at the sound of the bugle-call. Most trained domestic animals, dogs and oxen, and omnibus-and-car horses, seem to be machines almost pure and simple, undoubtingly, unhesitatingly doing from minute to minute the duties they have been taught, and giving no sign that the possibility of an alternative ever suggests itself to their mind."

Furthermore, habits are famously easier to develop when young. James (1983, pp. 116-117) cites Dr Carpenter here:

"The 'strength of early associations' is a fact so universally recognized, that the expression of it has become proverbial; and this precisely accords with the Physiological principle, that, during the period of growth and development, the formative activity of the Brain will be most amenable to directing influences."

And, given that this is the case, James (1983, p. 128), in an Aristotelian vein, entreats the young to:

"Seize the very first possible opportunity to act on every resolution you make, and on every emotional prompting you may experience in the direction of the habits you aspire to gain."

Adding:

"Could the young but realize how soon they will become mere walking bundles of habits, they would give more heed to their conduct while in the plastic state." (James, 1983, p. 130)

However, habits can mostly (at whatever age), although often with great effort, be modified. One can choose to train oneself to cease to act in certain previously habitual ways, and one can begin the adoption of new habits to overlay the old ones.

In sum, then, we have seen that habits essentially involve automatic behaviour that is the same (or relevantly similar) upon each occasion, and are acquired by drill, and hence non-human animals can acquire some habits.

Now, at this point, we might wonder whether in appealing to the notion of habit to capture Juliet's unconsidered behaviour, we are failing to recognise the complexity of her case. After all, the kind of habitual behaviour we have considered so far is what we might call *overt* behaviour. In particular, we have considered Ryle's example of walking, and James' of the evolutions of cavalry-horses. But Juliet's racist reactions take many forms and often do not involve overt behaviour. For example, Schwitzgebel (2010, p. 532) describes how:

"When she gazes out on class the first day of each term, she can't help but think that some students look brighter than others – and to her, the black students never look bright. When a black student makes an insightful comment or submits an excellent essay, she feels more surprise than she would were a white or Asian student to do so, even though her black students make insightful comments and submit excellent essays at the same rate as do the others."

Both of these examples are of what we might call *covert* reactions. They have to do with what Juliet is thinking and feeling, regardless of whether she evinces these attitudes in her overt behaviour. However, as Locke (1996, pp. 173-175) claims, it seems that such covert reactions can also be attributable to habit:

"Custom settles habits of thinking in the understanding, as well as of determining in the will, and of motions in the body; all which seems to be but trains of motion in the animal spirits, which once set agoing continue on in the same steps they have been used to, which by often treading are worn into a smooth path, and the motion in it becomes easy and as it were natural…this may serve to explain…[ideas] following one another in an habitual train, when once they are put into that tract, as well as it does to explain such motions of the body… Intellectual habits and defects this way contracted are not less frequent and powerful, though less observed."

He provides the following example of the habitual association of ideas:

"The death of a child, that was the daily delight of his mother's eyes, and joy of her soul, rends from her heart the whole comfort of her life, and gives her all the torment

imaginable…Till time has by disuse separated the sense of that enjoyment and its loss from the *idea* of the child returning to her memory, all representations, though never so reasonable, are in vain; and therefore some in whom the union between these *ideas* is never dissolved, spend their lives in mourning, and carry an incurable sorrow to their graves." (Locke, 1996, pp. 174-175)

However, we might wonder whether "habits of thinking" are simply habits of having or calling to mind certain beliefs. Perhaps Juliet habitually entertains racist beliefs, or makes racist judgements – for example, when she "can't help but think" that the black students do not look as bright as the others. But there is no need for us to admit that this is the case. Indeed, if Schwitzgebel – which on occasion he does – were to describe Juliet as *judging* in such circumstances, then this would be to beg the question against us. There is nothing to prevent us from describing the covert manifestation of Juliet's habit here as her habitually *entertaining the thought* that the black students look less intelligent.

This brings us to instinct, which we have proposed to appeal to in order to explain the woman's unconsidered behaviour in Precipice. James (1983, pp. 1036-1037) notes that "High places cause fear of a peculiarly sickening sort" and considers the "motor impulses" here to be of an "utterly blind instinctive character." But what is an instinct? Instincts seem to have several features in common with habits. For a start, they also prompt automatic behaviour that is the same (or relevantly similar) upon each occasion, and, like habits, they are possessed by non-human as well as human animals. Nonetheless, there is one key distinction between our concept of habit and that of instinct: namely, habits are thought of as *acquired* whereas instincts are considered to be *innate*. As James (1983, p. 1004) puts it:

"Instinct is usually defined as the faculty of acting in such a way as to produce certain ends, without foresight of the ends, and without previous education in the performance."

However, the notion of innateness is a controversial one (Griffiths, 2009, p. 1). At the beginning of the 20th century when James was writing it was common currency, but its status has since been thrown into doubt. Those who think the category a useful one, such as Konrad Lorenz (1965), often present results from the so-called *Deprivation Experiment* that was designed to test for innateness. In the Deprivation

Experiment, an animal is raised without learning a behaviour and then tested to see if it performs that behaviour when provided with an appropriate stimulus. If it performs the behaviour, then the behaviour is concluded to be innate. However, the Deprivation Experiment came under attack from Daniel Lehrman (1953), who pointed out that the experiment could not be used to conclude that a certain trait is innate *simpliciter*, but only that the factors controlled for in the experiment are not necessary for the development of that trait. In response, Lorenz (1965, p. 40) devised an alternative account of the innate/acquired distinction, according to which no trait is innate in itself, but:

"…certain parts of the information which underlie the adaptedness of the whole, and which can be ascertained by the deprivation experiment, are indeed innate" (cited in Griffiths, 2009, p. 4).

According to Lorenz, a trait is to be understood as innate insofar as its development is guided by *inherited information* as opposed to *environmental information*. But this response remains controversial. Further attempts have been made to account for the distinction between innate and acquired propensities in terms of *malleability*, the idea being that acquired propensities are typically more malleable than innate ones. But this idea is also disputed (Griffiths, 2009, p. 18). So, all-in-all, there is much disagreement as to whether the category of innateness is a legitimate one. Furthermore, of course, even if a legitimate use of the term could be agreed upon, it would be a further empirical question as to whether the propensity to experience fear at a great height is innate according to this use. So, our suggested folk psychological explanation of Precipice appears to be in trouble.

This brings us to Gendler's criticisms of the idea that we can simply offer a folk psychological explanation of cases such as Implicit Racist and Precipice, as opposed to relying upon her novel category of alief. Gendler proceeds by emphasising the similarities between the existing categories of habit and instinct by way of justification for grouping them both under the banner *alief*. Regarding the malleability of habits as opposed to instincts, she claims that:

"…it is not so clear either that this difference obtains in a relevant sense, or that malleability is what really matters." (Gendler, 2008c, p. 571)

It is unclear what she means by the latter claim, but we have seen that the idea that, roughly speaking, habits are malleable and instincts are not is controversial. To see this, we need only think of the "patient security guard" in Precipice. Why doesn't *he* behave as the woman does? Presumably because, through repeated exposure to the walkway, he has overcome his initial reaction to it. Indeed, Gendler claims that, to the extent that both instincts and habits are malleable, the process for manipulating them will be the same – she suggests exercises such as the refocusing of attention and the deployment of the imagination. Finally, regarding the etiology of habits as opposed to instincts, Gendler (2008c, p. 569) does not draw attention to the controversies surrounding the distinction between innate and acquired propensities we have discussed above, but, accepting the distinction, simply claims that:

"…it seems unmotivated…to distinguish propensities that result (directly) from the experiences and actions of a particular individual from those that result (indirectly) from the experiences and actions of her ancestors, merely on those grounds."

She draws an analogy with belief in this respect: we don't distinguish our beliefs on the basis of their origin. That is to say, some beliefs are derived *via* inference, some are perceptual and so on, but we nonetheless group them all together under the heading *belief.*

So, Gendler's strategy is to claim that the only really promising basis for distinguishing habit and instinct is their etiology, and that an appeal to this – by way of resistance to the idea that they are both aliefs – seems ill-motivated. In fact, she even says that she remains:

"…open to the possibility that there are distinct subspecies of alief: innate and habitual, perhaps – or controllable and uncontrollable. All that matters to my argument is that these subspecies be more similar to one another than they are to other candidate states." (Gendler, 2008c, p. 570)

However, we have seen that even an appeal to etiology in order to mark out habits from instincts is problematic. This seems to support Gendler's case for the introduction of the category alief all the more. So, does this mean that Gendler is right that it would be better to capture both Implicit Racist and Precipice by appeal to alief, as opposed to relying upon our existing folk psychological categories?

In a word: no. To take Juliet's case to begin with, despite the controversy surrounding the distinction between innate and acquired characteristics, it is clear that our concept of habit remains. Habits are essentially acquired *via* drill, and this is so whether we have any related innate propensities or not. And, as for Precipice, it is the fact that the woman is *afraid* that is really doing the explanatory work, regardless of whether this fear is instinctual in origin or not. It is because she is afraid that she experiences the phenomenology she does and she grabs hold of the security guard's arm, for example. So it seems that in spite of the controversy surrounding innateness that we have discussed, there are still robust folk psychological categories to which we may appeal in order to account for Implicit Racist and Precipice.

But then why not allow – as Gendler says that she would – that Juliet has a *habit alief* and that the woman in Precipice has a *fear alief*? After all, the cases are similar in the ways Gendler describes. The only way to answer this question is with another question: What would be the explanatory benefit of introducing alief into our explanations of such cases? As we have already seen, Gendler (2008c, p. 553) claims that doing so:

"…brings out the connection between a number of otherwise apparently discrepant issues, and renders unmysterious a number of otherwise perplexing phenomena."

But a competent user of folk psychology will already understand that both habits of thought and fears are similar in the kinds of ways Gendler describes – they do not need a new term to teach them this. As for the perplexing phenomena that are rendered unmysterious by the category alief: What is left mysterious when we claim that Juliet has a recalcitrant habit and that the woman in Precipice is afraid?

To put the point as clearly as possible, given that Gendler is recommending that we augment our folk psychology with the introduction of a novel concept, the burden of proof as to its utility rests firmly with her, and it is this that she is yet to establish. Indeed, one might think, as Gendler (2008c, p. 557) observes in a footnote, that the very fact that we have omitted a folk psychological notion of alief up until this point is itself revealing. Gendler (2008c, p. 557) admits that she has "no convincing response to this worry"; she simply stresses that she does not propose to replace existing concepts such as instinct and habit with that of alief, but simply to augment

our conceptual resources. What remains to be seen is whether alief is a necessary augmentation.

1.6 The Rational and the Arational

Despite our disagreement with Gendler over the utility of the category of alief, our proposed explanation of cases such as the Implicit Racist and Precipice reveals a key point of agreement with her approach. Gendler (2008c, p. 565) explains that the central issue between her and those such as Schwitzgebel is "the role that belief needs to play in our cognitive repertoire", and whether or not it should be "intimately connected with notions like knowledge and rationality". She claims that it should be:

"…whatever belief is – it is normatively governed by the following constraint: belief aims to 'track truth' in the sense that belief is subject to immediate revision in the face of changes in our all-things-considered evidence… Beliefs change in response to changes in evidence; aliefs change is response to changes in habit." (Gendler, 2008c, pp. 565-566)

So, Gendler maintains that we ought to describe Juliet as believing in accordance with her reasons that the races are intellectually equal, whilst alieving arationally that this is not the case. Similarly, the woman in Precipice believes in accordance with her reasons that she is safe, whilst alieving arationally that this is not the case. Whereas Schwitzgebel treats all of Juliet and the woman in Precipice's behaviour as on a par with respect to what they believe (his egalitarianism), Gendler identifies their reflective behaviour as pertinent to what they believe, as it evinces what they understand their reasons to be, and their unreflective behaviour as displaying their arational associations, and so pertinent to what they alieve.

Despite rejecting Gendler's introduction of the concept alief, we are in broad agreement that in cases such as Implicit Racist and Precipice, it is the person's considered behaviour that is pertinent to what they believe, as it is importantly connected to what they take their reasons to be. However, Schwitzgebel (2010, pp. 540-541) objects that to maintain that there is a privileged relationship between belief and reasons:

"…artificially hives off our rational and thoughtful responses from our habitual, automatic, and associative ones. In asserting that only the first are pertinent to

belief… Gendler attempt[s] to separate what is really an inseparable mix. People judge in part automatically, associatively, and arationally, and they often show high intelligence in their habits and their unreflective, spontaneous responses. Words spring from our mouths seemingly unbidden, sometimes to our regret and sometimes reflecting a wit and intelligence that surprises even the speaker, as in the skillful flow of repartee…. Teachers of formal logic and critical thinking swiftly and spontaneously recognize affirmations of the consequent in student homework, probably using pattern-recognition abilities not so different from those used to respond to other sorts of patterns; and in conversation, too, we detect fallacies and falsities using, it seems likely, a suite of automatic, habitual, and associative abilities, as well as more overtly rational ones."

Gendler (2008c, p. 566) admits that the kind of division she draws between the rational and the arational is an idealisation. As we will begin to explore in detail in the next chapter, in reality, there are not two neatly contained regions of the mind, one operating in perfect accordance with reason and pertaining to belief, and another at a wholly arational level, totally impervious to reason and pertaining to other mental states. But we will argue that this does not undermine the idea that there is a privileged relationship between belief and reasons.

Conclusion

Before we move on to begin to mount a defence of this claim, it will be helpful for us to summarise what we have achieved so far. We began by describing two cases in which it is difficult to know what to say the person in question believes: Implicit Racist and Precipice. We then outlined the four kinds of approach one could take to such cases if one assumed belief to be a discrete matter. We chose to explore examples of two of those options, as representative of the two poles of the debate over the nature of the relationship between belief and reasons. The first was an approach invited by dispositionalism about belief, outlined by Schwitzgebel (2001b, 2010), which is in keeping with the idea that there is no significant relationship between belief and reasons. Given Schwitzgebel's egalitarianism – his assumption that every disposition in the dispositional stereotype of a particular belief is on a par with every other – he thinks that someone who possesses some of these dispositions and not others, such as Juliet and the woman in Precipice, is in-between believing.

We began to see what is wrong with this kind of approach when we considered how it failed to explain or predict the *pattern* in the behaviour of both Juliet and the woman in Precipice. In both cases, their considered behaviour all hangs together, as does all of their unconsidered behaviour. We suggested that a folk psychological description of each case which, roughly speaking, locates belief with reasons could capture these patterns. We then considered Gendler's (2008b, 2008c) alternative approach to these cases, which also emphasises the relationship between belief and reasons. However, we rejected Gendler's claim that the addition of the concept of alief to our existing stock of folk psychological concepts is a beneficial one. Nonetheless, we share Gendler's conviction that there is a significant relationship between belief and reasons, and it is this idea that we will now go on to explore and defend throughout the rest of this thesis. We will see that if we keep belief in its proper place with reasons, and rely upon our rich stock of other folk psychological concepts for further explanatory power, we do not have to twist the concept of belief out of shape as Schwitzgebel does, trying to make belief – in combination with desire – account for all of our behaviour.

2. Transparency, Heterogeneity, and Aiming at Truth

In the previous chapter, we saw that conceiving of belief and reasons as intimately related to one another enables us to capture the patterns in the behaviour of both Juliet and the woman in Precipice. The proposal was that their considered behaviour is in some sense revelatory of what they understand their reasons to be and hence of what they believe. However, as yet we have said little about the nature of this relationship between belief and reasons or in defence of the claim that it is an important one – other than indicating that the idea can help us capture these patterns. Moreover, we have acknowledged that the relationship is not by any means a straightforward one. In this chapter we will begin to explore the connection between belief and reasons in greater detail and defend the claim regarding its significance, in spite of its complexity. We will maintain that what is revealed by this relationship is that it is a constitutive fact about belief that it *aims at truth*, as Williams (1973) famously put it. In so doing, we will present a novel interpretation of the manner in which this claim is to be understood, which surmounts previously unresolved difficulties.

We will begin by drawing some crucial distinctions when it comes to understanding the notion of a reason, which we will need to proceed. We will then highlight an important feature of the manner in which belief and epistemic reasons appear to interact, namely, what Shah (2003, 2005) has called the *transparency* of doxastic deliberation. This is the idea that when one consciously considers *whether to believe that p*, one experiences this question as (non-inferentially) immediate to the question as to *whether p*. This claim about the phenomenology of doxastic deliberation seems to suggest that belief and reasons are importantly related. However, although from a first-person perspective, it does seem as though one simply considers one's reasons when considering what to believe, nonetheless, when we step outside of such a perspective and look on at the beliefs of others and even our own beliefs from a third-person perspective, we realise that factors other than truth do influence belief. It is commonplace that what one believes is often influenced by factors such as what one *desires* to be true, for example, although it seems that this kind of influence can only take place outside of one's conscious awareness. We will call the fact that what one believes often depends upon factors other than one's reasons *heterogeneity*, to reflect the fact that there can be different kinds of influence upon belief.

Heterogeneity seems to add weight to the position of those such as Schwitzgebel, who deny that there is a relationship of any particular significance between belief and reasons.

But in this chapter we will argue that this is an illusion: in spite of heterogeneity, there is nonetheless a significant connection between belief and reasons, which reveals that beliefs aim at truth. How the claim that beliefs aim at truth is to be understood is of course controversial. We will explore two main conceptions of it. First, what we will call the *regulation* conception, according to which it is constitutive of belief that it is regulated (formed, upheld, revised, and extinguished[8]) for truth; secondly, what we will call the *correctness* conception, according to which the belief that *p* is *correct* if and only if *p* is true. David Velleman (2000) affirms both of these conceptions of the claim that beliefs aim at truth and maintains that the latter can be derived from the former. Roughly speaking, his idea is that if beliefs are regulated for truth, then truth is in some sense the goal of belief. Accordingly, the normative force that truth has for belief is the normative force that a goal has for whomever's goal it is. Because beliefs are regulated for the goal of truth, a belief is correct if and only if it is true. However, we will argue that although beliefs do aim at truth in both the regulation and correctness senses, the correctness conception of this claim is the most fundamental. Belief is at base a normative concept about which, as John McDowell (1994, p. 60) claims, "the question of one's entitlement… can always be raised". From this, we can then derive the claim that beliefs are regulated for truth in the sense that if and when they are consciously regulated, they are so-regulated exclusively for truth. Transparency is thus explained by the fact that when one considers *whether to believe that p,* one understands that belief is regulated for truth in this way because one understands that beliefs are correct if and only if their contents are true. The claim that correctness can explain transparency has recently been subject to criticism on the grounds that our recognition that we ought to believe the truth could not motivate us solely to consider whether *p* (Owens (2003), Asbjørn Steglich-Petersen (2006b), and Conor McHugh (2011c)). In the course of providing

---

[8] Shah and Velleman (2005, p. 498) understand regulation to involve the manner in which one's beliefs are "formed, revised, and extinguished" but we have added *upheld* to the list for the sake of completion. It should be noted that talk of *regulation* might to taken to imply a conception of our control over our own beliefs as involving acting *upon* them. In chapter four, following Matthew Boyle (forthcoming a), we will argue that our control over our beliefs does not conform to such a model. For now, let us simply note that talk of regulation is intended to be understood as neutral regarding the manner in which agents exercise control over their beliefs.

our account, we will therefore respond to such claims, arguing that they rest on both a failure to distinguish epistemic and practical norms, and a failure to appreciate the inherent normativity of the deliberative standpoint. We close by reviewing our initial assessments of Implicit Racist and Precipice and explain how the conception of belief we have begun to develop supports these assessments.

2.1 What is a Reason?

So far, we have used the term *reason* very casually, but in order to come to understand the significance of reasons for belief we must of course have a slightly closer look at how to understand the term. In particular, there are three important points about how we shall be using the term to be made clear. First, so far we have talked about reasons for believing *p* to be true. More precisely, these are what we shall call *epistemic* reasons. By way of contrast, we have mentioned factors other than reasons for believing *p* to be true. Now, these factors could often be reasons themselves – such as wanting to believe that *p*, for example – but this is a reason of another type: it is what we shall call a *non-epistemic* reason. So, we can distinguish between two types of reason along this dimension, which, of course, correspond to the traditional distinction between *theoretical* and *practical* rationality.

Secondly, we shall distinguish the term *reason* along another dimension in terms of what we shall call *subjective* and *objective* reasons. A subjective reason is something one takes to be a reason for something, whether or not one is correct to do so, whereas an objective reason is a reason for something else, whether or not one takes it to be. So, one's subjective reasons for believing that *p* can also be objective reasons for believing that *p*. However, one's subjective reasons can depart from how things are objectively (in which case, as we will later see, any belief formed on the basis of one's subjective reasons will count as irrational). Although in what follows we will often prefix *reason* with either *epistemic* or *non-epistemic* and *subjective* or *objective*, if we refer simply to a *reason*, it is to be understood as a subjective epistemic reason.

Finally, it is important to note here that we will be operating throughout with a fairly permissive conception of a reason. For example, one's reason for believing that there is a wine glass in front of oneself can simply be one's seeing that this is so.

Furthermore, we shall allow for unconscious subjective reasons as well as conscious ones.

2.2 The Transparency of Doxastic Deliberation

A key feature of the relationship between belief and reasons that appears to suggest that it is a significant one for understanding the nature of belief is what Shah (2003, p. 448) calls the *transparency* of doxastic deliberation. *Transparency* in the present sense needs to be carefully distinguished from the phenomenon also known as *transparency* in the self-knowledge literature. In this literature, the claim that our knowledge of our own beliefs is transparent to the states of affairs those beliefs refer to is often understood along the lines of Gareth Evans' famous claim (1982, p. 225) that:

"…in making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward – upon the world. If someone asks me 'Do you think there is going to be a third world war?', I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' I get myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*…If a judging subject applies this procedure, then necessarily he will gain knowledge of one of his own mental states: even the most determined sceptic cannot find here a gap in which to insert his knife."

The phenomenon that Shah refers to as the transparency of doxastic deliberation, however, is different. Roughly speaking, Evans is claiming that one knows *whether one believes that p* by considering *whether p*. In contrast, Shah claims that in attempting to answer the question *whether to believe that p* one experiences this as a question which non-inferentially gives way to a question as to *whether p*.

So, doxastic deliberation is understood by Shah to be framed by the question *whether to believe that p*. However, he is keen to emphasise that one can be in the context of doxastic deliberation without oneself *explicitly* posing this question:

"…all that is required is that the question be in the background of… [one's] reasoning, guiding… [one's] deliberation." (Shah, 2003, p. 467)

This characterisation of the phenomenology of doxastic deliberation is a plausible one. So, how can we explain it? It seems that the sort of explanation we require is likely to point to some significant connection between the concepts of belief and truth – for how else could we explain the transparency of the question *whether to believe that p* to *whether p* other than by citing such a connection? For this reason, following Velleman's (2000) lead, we can turn to Williams' (1973) famous claim that *beliefs aim at truth*. Thus, we will now consider whether there is a construal of this claim that is able to capture the transparency of doxastic deliberation.

2.3 Belief as Aiming at Truth

An obvious place to begin to try to understand this claim is to turn to Williams' original statement of it (1973). Williams explains that he means three things by the claim. The first is that:

"…truth and falsehood are a dimension of an assessment of beliefs as opposed to many other psychological states or dispositions… If a man recognises that what he has been believing is false, he thereby abandons the belief he had." (Williams, 1973, p. 137)

Both of these related ideas seem plausible and we will return to them in due course. Before we do so, though, let us examine Williams' other two elaborations of his claim that beliefs aim at truth.

The second idea Williams thinks we should understand by the claim is that: to believe that *p* is to believe that *p* is true. Now, Velleman (2000, p. 247) has made a convincing case against such an interpretation of the claim that beliefs aim at truth, explaining that:

"…every instance of believing is an instance of believing something to be true, and this relation to the truth is sometimes confused with truth-directedness. But in bearing this particular relation to the truth, belief is just like any other propositional attitude, since wishing entails wishing something to be true, hoping entails hoping something to be true, desiring entails desiring something to be true, and so on. Hence the fact that believing entails believing-true doesn't set belief apart from other propositional attitudes, as truth-directedness is supposed to do."

So, if we thought of the claim that beliefs aim at truth as to be understood as the claim that believing involves believing-true, we would be forced to admit that the claim says nothing distinctive about belief, because every propositional attitude involves ∪-ing-true. And, given that our interest in the claim that beliefs aim at truth is an interest in a claim that purports to distinguish belief from other mental entities, we had better not understand the claim to mean that believing involves believing-true.

Williams' (1973, p. 137) third suggestion for understanding the claim is that: "to say 'I believe that *p*' itself carries, in general, a claim that *p* is true." He thinks that it is important to recognise that this is so in order to help us to understand why Moore's Paradox is paradoxical. That is, Williams' third suggestion enables us to understand why the claim *I believe that it is raining but it is not raining* is paradoxical. This is because saying *I believe that it is raining* carries an implied claim that *p* is true. So, in *saying I believe that it is raining but it is not raining*, one implies both that it is true that it is raining, and that it is false that it is raining. Now, it would be difficult to disagree with Williams here, but it seems that this claim about our speech acts fails to hit rock bottom when it comes to understanding what it means to say that beliefs aim at truth.[9] Indeed, in saying that asserting *I believe that p* means something a bit like *p is true* it seems as if Williams could be simply re-asserting the phenomenon we seek to understand with the claim that beliefs aim at truth. Even if this is not so, that *I believe that p* means something a bit like *p is true* is certainly not a wholly convincing explication of the idea that beliefs aim at truth.

Rather than directly investigating either Williams' second or third suggestions, we will turn instead to the two main interpretations of the claim that beliefs aim at truth that have emerged from contemporary debate of the matter, both of which have their roots in Williams' first suggestion. These are what we will call the *regulation* conception of the claim and the *correctness* conception.

---

[9] This is not to imply that Williams thought of his claim about our speech acts involving the concept belief as the ultimate analysis of the claim that beliefs aim at truth. Clearly, he did not think it was: he lists it as the third of three analyses, and describes it simply as "closely connected" with the previous two (Williams, 1973, p. 137). The present point is simply that this third claim of Williams' ought properly to be understood as a consequence of the claim that beliefs aim at truth, rather than an analysis of it.

The regulation conception is well understood as an expansion of Williams' (1973, p. 137) claim that "[i]f a man recognises that what he has been believing is false, he thereby abandons the belief he had". According to the regulation conception, the claim that beliefs aim at truth is a claim about the manner in which one's beliefs are regulated – that is, the manner in which one's beliefs are formed, upheld, revised, and extinguished.[10] On the regulation interpretation, the claim that beliefs aim at truth amounts to the idea that it is constitutive of belief that it is regulated for truth. (Proponents of this view include Velleman (2000), Shah (2003), Shah and Velleman (2005), Steglich-Petersen (2006b), Daniel Whiting (2012).)

Meanwhile, the correctness conception of the claim that beliefs aim at truth is fruitfully understood as an explication of Williams' (1973, p. 137) claim that "truth and falsehood are a dimension of an assessment of beliefs as opposed to many other psychological states or dispositions." According to the correctness interpretation, the claim that beliefs aim at truth is to be understood as a claim about belief's *standard of correctness*: that the belief that *p* is *correct* if and only if *p*. Precisely how this idea is to be understood is controversial, as we shall see in chapter four, but many have shared this conviction in one form or another (e.g. Velleman (2000) Paul Boghossian (2003), Shah (2003), Shah and Velleman (2005), McHugh (2012), Ralph Wedgwood (2002), Whiting (2010)).

As the presence of many of these authors in both of the above lists makes clear, the regulation and correctness conceptions of the claim that beliefs aim at truth are not mutually exclusive. One can maintain that both the regulation and the correctness claim are constitutive of belief. In this chapter, we will explore both of these conceptions and consider whether either or both are well-placed to capture the transparency of doxastic deliberation. Before we begin to undertake this task, however, it will be important, briefly, to focus our attention on two caveats regarding how the claim that beliefs aim at truth is to be interpreted. These are as follows.

---

[10] It is interesting to note that it may be difficult to distinguish between the revision and the extinguishing of a belief on certain occasions. For example, imagine that *S* believes that the number of planets in our solar system is nine, but is subsequently informed that Pluto is not (strictly speaking) a planet. Should we count *S* as extinguishing her belief that there are nine planets in our solar system and replacing it with the belief that there are eight or as simply partially revising the contents of her existing belief from nine to eight? (It is not obvious how we should go about answering this question.)

First, we are not conceiving of the claim that beliefs aim at truth as pertaining to an activity one might call *inquiry*. We can think of inquiry as the process of gathering evidence on a certain topic. The process of belief regulation we will be interested in is that which occurs once the evidence is already in – not the search for further evidence.

Secondly, the claim that beliefs aim at truth is to be understood as pertaining to beliefs *individually*, not as some comment on our global epistemic projects. It is not the claim that our overarching epistemic goal is/ought to be to believe as many truths as possible, for example. As Velleman (2000, p. 251) points out, this would of course be (practically) irrational, "since the world is teeming with truths, most of which are too trivial to be worth believing." It is not even to say that we do/ought to aim at the practically rational goal of believing as many interesting, or useful truths as possible (Velleman, 2000), nor at maximising "the net balance of truth in… [our] beliefs", as Peter Railton (1994, p. 73) puts it. Aiming at these goals would allow us to believe falsehoods if it better enabled us to achieve our end (Railton, 1994). Rather, the idea that beliefs aim at truth is supposed to individuate beliefs and, as Velleman (2000, p. 252) says:

"…one's acceptance of a proposition can amount to a belief without being part of any global epistemological project of accumulating true beliefs."

2.4 The Heterogeneity of Belief Regulation

So far then, we have outlined Shah's (2003) plausible characterisation of the phenomenology of doxastic deliberation in terms of transparency, and we have noted that it seems as though the idea that beliefs aim at truth could be invoked in order to explain such transparency. However, we are yet to see which interpretation of this claim would best explain transparency. Furthermore, there is a potential barrier to thinking of belief as aiming at truth that it would be natural for one such as Schwitzgebel who sought to undermine the idea that belief and reasons are importantly connected to emphasise. That is, although in the context of doxastic deliberation it appears that the question *whether to believe that p* is transparent to the question *whether p*, it is a key aspect of folk psychological wisdom that outside of the believer's conscious awareness, belief regulation is often influenced by non-epistemic reasons. Factors such as whether the person *wants* to believe that *p* often

influence whether or not they do, for example. This is what we have called the *heterogeneity* of belief regulation. Can either the regulation or the correctness conception of the claim that beliefs aim at truth capture *both* the transparency of doxastic deliberation *and* the heterogeneity of belief regulation?

2.5 Transparency and Regulation

We will begin by considering whether the regulation conception can capture transparency. To this end, we will start by presenting Velleman's (2000) classic account of the claim that beliefs aim at truth, according to which the fundamental constitutive feature of belief is that it is regulated for truth. We will then consider Shah's (2003) criticism of this account according to which it is unable, as it stands, to capture the transparency of doxastic deliberation. As we will see, if this criticism is successful it will follow that we cannot provide a satisfactory explanation of transparency solely in terms of regulation. However, Steglich-Petersen (2006b) maintains that we can explain transparency solely in terms of regulation and provides an account designed to avoid Shah's criticisms. We will argue that such an account is indeed possible, but that Steglich-Petersen's constructive proposal fails. In its place, we will offer an alternative regulation-based explanation of transparency.

What we might call the *early* Velleman – Velleman (2000) – maintains that beliefs aim at truth in *both* the regulation and the correctness senses detailed above. However, he thinks that the correctness condition is *derivable* from the regulation condition:

"Let me be clearer about the relation between the constitutive aim of belief and the norm that applies to belief in light of that aim. To say that belief aims at the truth is not simply to re-express the norm stipulating that a belief must be true in order to be correct; rather, it is to point out a fact about belief that generates this norm for its correctness." (Velleman, 2000, pp. 16-17)

In other words, Velleman thinks that beliefs are the kind of entities that are regulated for truth, and that this entails that they have truth as their norm of correctness. Roughly speaking, the idea is that if beliefs are regulated for truth, then truth is in some sense the goal of belief. Accordingly, the normative force that truth has for belief is the normative force that a goal has for whomever's goal it is. Because

beliefs are regulated for the goal of truth, the belief that *p* is correct if and only if *p*. But what, exactly, does Velleman understand by the idea that belief is regulated for truth and how is this to account for both transparency and heterogeneity?

Velleman (2000) does not discuss transparency explicitly, but according to his account, transparency would be explained by the fact that the doxastic deliberator understands the concept of belief they employ to be that of an entity that is regulated for truth. Now, Velleman (2000, p. 253) recognises that not all beliefs are regulated intentionally in the manner of the doxastic deliberator's; beliefs can also be unconsciously regulated for truth:

"Suppose that one part of the person – call it a cognitive system – regulates some of his cognitions in ways designed to ensure that they are true, by forming, revising, and extinguishing them in response to evidence and argument. Regulating these cognitions for truth may be a function for which the system was designed by natural selection, or by education and training, or by a combination of the two. In any case, the system carries out this function more or less automatically without relying on the subject's intentions for initiative or guidance."

He maintains that many (perhaps the majority) of our beliefs are regulated for truth in this manner. Secondly, Velleman claims that he:

"…doesn't require belief to be governed by truth-seeking mechanisms *alone*. There are probably psychological mechanisms that cause, and are designed to cause, beliefs that happen to diverge from the truth. Evolution or education may have given us dispositions to err on the side of caution in perceiving predators, to overestimate our own popularity, and so on. *But my thesis is not that belief is completely shielded from mechanisms that tend to make it false; my thesis is that belief is necessarily subject to mechanisms designed to make it true.*" (Velleman, 2000, p. 254) (italics added)

So, Velleman claims that belief is *weakly* regulated for truth, in the sense that although belief is always regulated for truth to some extent, it is also often influenced by factors other than truth, such as what the believer desires to believe, for example. We will refer to this weak regulation claim as (Rw). By acknowledging (Rw), Velleman attempts to accommodate heterogeneity as well as transparency.

Now, at this point, one might wonder what is to stop Schwitzgebel from objecting that not only is heterogeneity true, but that unconsciously regulated beliefs can be regulated *exclusively* for factors other than truth. Velleman's (2000) response would be that we need to think of regulation as *constitutive* of belief, not merely incidental to it, in order to explain transparency. On his picture, it is simply in virtue of the doxastic deliberator understanding the concept of belief as regulated for truth that we can explain transparency. Thus, if unconsciously regulated states are to count as beliefs, they too must be regulated for truth – although this can only be in the sense of (Rw), because of heterogeneity.

However, Shah (2003) objects to Velleman's (2000) conception of belief as follows. Shah (2003, p. 462) agrees with Velleman that it is:

"…entrenched in common sense that reason does not always have decisive doxastic influence."

So, he thinks that we must accept (Rw) as opposed to (Rs) – the claim that belief is *strongly* regulated for truth (regulated *solely* for truth). However, given that in the context of doxastic deliberation, one is employing (albeit sometimes implicitly) the concept of belief, one is reliant upon one's grasp of the concept of belief. And, if one *knows* that beliefs are often regulated for matters other than truth, what is to prevent one from thinking the following?

Well, I know the evidence seems to point towards the fact that if I prod a little deeper I'll discover that p, but I really couldn't cope if I discovered it to be the case, so I'll believe, in accordance with what I'd like to believe, that ~p.

Although from the outside we can see that what people believe is often decisively influenced by factors other than the truth in this manner – even in the case of beliefs that are the up-shot of doxastic deliberation – the person in the context of doxastic deliberation *cannot* think in this way: doxastic deliberation is transparent.

Furthermore, Shah argues that even if Velleman were to alter (Rw) to (Rs), he would not be able to account for the transparency of doxastic deliberation. Shah (2003, p. 469) thinks that this is because:

"The focus on truth… [for Velleman] is of *instrumental* value in the achievement of the goal of belief; therefore the focus on truth is *mediated* by an inference that it would be conducive to my goal to focus on truth."

And, Shah claims that the phenomenology of doxastic deliberation involves consideration of *whether to believe that p immediately* giving way to consideration of *whether p*: no inference is involved.

Thus, if Shah's criticisms go through, then no formulation of the regulation claim – whether in terms of strong or weak regulation – will be capable of capturing transparency. In the following, we will consider whether Shah gives us convincing grounds to embrace such a conclusion.

Let us deal with the second of Shah's (2003) criticisms of Velleman (2000) first. Shah thinks that, according to (Rs), in the context of doxastic deliberation, the subject considers *whether to believe that p* and thinks:

Well, beliefs are the kind of thing which are regulated solely for truth, so I can only consider whether or not p is true in order to decide whether or not to believe that p.

In reality, however, an account in terms of (Rs) could easily resist such caricature. The proponent of such an account could hold that it is simply in virtue of understanding the concept of belief that the doxastic deliberator understands that (Rs) holds. No inference has been shown to be involved: just a simple grasping of the concept *belief*.

Nonetheless, both Shah (2003) and Velleman (2000) are agreed that (Rs) cannot be considered to be constitutive of belief because of heterogeneity. So, does this mean that Shah (2003) is right that we cannot account for transparency solely in terms of regulation? Steglich-Petersen (2006b) maintains that this is not the case: we can explain transparency solely by appeal to regulation. Let us now turn to his attempt to do so.

Steglich-Petersen claims that we need to recognise that there are two main concepts of belief at work in our thinking. He distinguishes between what he calls the *empirical concept* of belief, according to which a mental state counts as a belief even if it is only weakly regulated for truth, and the *core concept* of belief, according to

which beliefs are strongly regulated for truth (Steglich-Petersen, 2006b, p. 511).[11] The latter is the concept we employ in the context of doxastic deliberation, and:

"…while we cannot explain the role of the concept of belief in doxastic deliberation in terms of weak truth-regulation, we can explain why weakly truth-regulated states and processes nevertheless count as 'beliefs' in terms of the primary deliberative concept of belief." (Steglich-Petersen, 2006b, p. 511)

According to Steglich-Petersen, weakly truth-regulated states can count as beliefs in two different ways. First, because even though they are only weakly regulated for truth, the *believer* is nonetheless consciously intending or aiming (we will use these two terms interchangeably) to accept the truth and only the truth. In order to understand this idea, we must remind ourselves that we are trying to explicate the claim that *beliefs* aim at truth. So, it is important to distinguish talk of *believers* as aiming or intending and *beliefs* themselves as so-doing. Believers can literally aim or intend; beliefs cannot. Thus it is the latter, rather than the former notion that requires explication. Steglich-Petersen's idea is that we can understand the claim that *beliefs* that are only weakly regulated for truth nonetheless aim at truth insofar as this is the *believer*'s aim.

Imagine a darts player who, *qua* darts player, is aiming at the bull's eye on the dart board. Simply because she fails to hit the bull's eye because of a tiny twitch in her hand as she lets go of the dart does not mean that she was in fact torn between aiming at the left hand side of the board and the bull's eye. Similarly, in being influenced (unbeknownst to them) by what they desire to believe, this does not mean that the believer is not solely aiming at truth. It is simply that they are not doing very well at it on this occasion. Now, of course, this analogy is not perfect because an involuntary twitch of the hand is different in kind to the unconscious influence of a desire. Nonetheless, what both of these phenomena have in common is that they are non-intentional, and this is all we need for the analogy to do its work.

But what about beliefs that are formed at the unconscious level? According to Steglich-Petersen, these beliefs represent a third division in our concept of belief,

---

[11] When formulating the truth aim, Steglich-Petersen (2006b) at times uses *only if*, but mostly *if and only if*. He seems to intend the latter, so this is the formulation we will attribute to him. Our discussion of his position is unaffected by this decision.

between the two types of beliefs that, unlike those falling under the core concept, are thought of as weakly regulated for truth. There are those beliefs that are weakly regulated for truth in which the believer nonetheless intends to believe the truth, and there are those which are weakly regulated for truth at the unconscious level. He claims that:

"Such blind processes count as beliefs… because they share their essential descriptive characteristic of weak truth-regulation with processes that have that characteristic in virtue of how they are related to the intentional aims of a believer." (Steglich-Petersen, 2006b, p. 515)

So, beliefs that are only weakly regulated for truth but are nonetheless intentional are to be understood as beliefs in terms of our core concept of belief according to which we intend to believe the truth. And, beliefs that are only weakly regulated for truth and are not even intentional inherit *their* status as beliefs from weakly truth-regulated beliefs which were the product of an intention. How does this latter move work? Steglich-Petersen (2006b, p. 514) claims that although the empirical concept of belief is to be understood in terms of its intentional counterpart, it still has some "independent empirical content". That is, he thinks that it nonetheless contains the idea that a mental entity must be at least weakly regulated for truth to count as a belief. Unconsciously regulated beliefs share:

"…precisely the descriptive characteristic of being to some degree conducive to the *hypothetical aim* of someone intending to form a belief in the primary strong sense." (Steglich-Petersen, 2006b, p. 515) (italics added)

So, Steglich-Petersen concludes that the regulation condition alone can capture the transparency of doxastic deliberation and explain the status of weakly truth-regulated mental entities as beliefs. So how does he propose to explain transparency? His claim is that in the context of doxastic deliberation, the deliberator employs the core concept of belief as strongly regulated for truth and hence understands that it is:

"…a simple conceptual truth that intending to form a belief about *p just is* intending, or aiming at, accepting *p* if and only if *p* is the case." (Steglich-Petersen, 2006b, p. 516)

On this picture, transparency is the product of the doxastic deliberator (necessarily) appealing to one concept of belief amongst others, and understanding their own situation in light of this concept. But precisely because of this feature of Steglich-Petersen's account, it seems that Shah's question could simply be reposed. *Why*, rather than the core concept, doesn't the doxastic deliberator employ the empirical concept of belief Steglich-Petersen defines? After all, they possess this concept as well. Now, Steglich-Petersen correctly insists that the doxastic deliberator does not think as follows:

I know that this state will count as a belief even though it's only weakly regulated for truth, so I'll adopt it instead of a state strongly regulated for truth.

However, what we cannot do is to explain why this is in a non-question-begging way. If we accept that there are two main concepts of belief – the core and empirical – the question is why the doxastic deliberator employs one rather than the other. Simply stating that they do does not seem to explain anything.

Nevertheless, the failure of Steglich-Petersen's account does not signal the end for regulation-based accounts. In what follows we will present an alternative picture, which avoids the pitfalls facing Steglich-Petersen's account by appealing to a *unitary* concept of belief.

2.6 An Alternative Picture

It seems that, in spite of heterogeneity, we can capture transparency solely by appealing to strong regulation if we do so as follows: we maintain that in possessing the concept of belief, the doxastic deliberator understands that, however the belief that *p* is in fact regulated, if and when it is consciously regulated, this must be solely for truth. That is, although they know that unconsciously regulated beliefs are often regulated for factors other than truth, they also know that in so far as it is to be regulated at the *conscious* level, it can only be so-regulated for truth. So, the doxastic deliberator cannot think: *although it's only weakly regulated for truth, this state would count as a belief.* This is because even though they know that if regulated unconsciously the state in question may well count as a belief, *ex hypothesi* they are consciously regulating it. And, in so far as they understand the concept of belief, they

understand that such conscious regulation can only be for truth. So, we have arrived at a conception of belief according to which:

(R): It is constitutive of belief that *if* it is consciously regulated, it is so-regulated solely for truth.

Unlike Steglich-Petersen's account, this provides us with a unified conception of belief. On the present account, it is the same concept of belief that we appeal to in classifying unconsciously regulated beliefs as is appealed to by the doxastic deliberator. On Steglich-Petersen's picture, it is as if the doxastic deliberator selects from a number of concepts of belief, and Shah's problematic can thus re-emerge through the question as to why they must employ one concept rather than another. The present account, however, does not create the same space for Shah's question, because on our account, the doxastic deliberator is not employing one concept to the exclusion of another. Both from a first- and third-person perspective, belief is conceived in terms of the way in which it is consciously regulated, regardless of how it is unconsciously regulated.

Now, before we proceed, a number of points must be emphasised about the commitments and status of (R). First, we are not claiming that it is only by being consciously regulated for truth that beliefs may feature in consciousness. (R) is quite compatible with the everyday occurrence of beliefs being present in consciousness, without being regulated. A friend asks me to remind him when the Battle of Hastings took place, and I automatically respond, *1066*. If the subject of conversation changes without me being asked to provide any grounds for my answer, then the belief that the battle took place in 1066 has been consciously entertained without having been consciously regulated. By the lights of (R), this does nothing to threaten the idea that I believe that the Battle of Hastings took place in 1066. Secondly, (R) is compatible with a relatively permissive conception of conscious truth-regulation. If, to pursue our present example, I go on to ask myself why I believe that the battle took place in 1066, then something as minimal as *I'm pretty sure I learned that at school* could suffice as a form of conscious truth-regulation, given the background assumption on my part that dates learned at school are generally correct. Thus, some minimal degree of reflection on the status of the sources of one's beliefs may suffice for conscious truth-regulation: in the present case, one needn't go back to a text book. Finally, it

should be noticed that (R) involves no reference to the believer themselves aiming at truth. This is intentional: the question of whether (R) involves the possession of such an aim on the part of the believer will be considered at length in the next chapter.

Now, given that it claims to pick out a constitutive feature of belief, the truth of (R) would be incompatible with our viewing non-human animals and human infants as capable of holding beliefs, supposing that they are incapable of conscious truth-regulation. Nonetheless, there are a number of obvious strategies that the proponent of (R) might adopt in order to make our practices of attributing beliefs to such creatures intelligible: for example, it might be claimed that such attributions involve an analogical extension from our own situation, it which case it would perhaps be more accurate to characterise non-human animals and infants as possessing proto-beliefs. Given the possibility of such strategies, it would be implausible to claim that (R) is made untenable in the face of animals and infants: whilst all parties can agree that a satisfactory account of belief must be able to explain our practice of attributing beliefs to such creatures, it would be tendentious to insist in advance of enquiry that a satisfactory account of belief must show that animals and infants do in fact possess beliefs.

So far, in our explanation as to why unconsciously regulated states count as beliefs, we have appealed to a counterfactual pertaining to them: *if* they come to be consciously regulated, this will be solely for truth. Now, one might wonder whether in addition to being subject to this counterfactual, these unconscious states must also be weakly regulated for truth to count as beliefs. In fact, it seems that the claim that (Rw) is constitutive of belief is intuitively doubtful, as the following thought-experiment brings out. Recall that we have characterised regulation in general as involving formation, upholding, revision, and extinction. In the case of weak regulation, when we prise apart these different strands we can see that it is not obvious that it is constitutive of belief. In particular, whatever we think about upholding, revision, and extinction, it is not clear that a mental state need be *formed* for truth in order for it to be intuitively classifiable as a belief. To see this, let us imagine a case in which an evil demon implants a mental state with the content *p* into *S*'s mind on a whim. Now it seems that we are still prepared to think of this state as a belief if, despite coming into it in this way, *S* nonetheless currently thinks that she has sufficient epistemic reason in favour of *p*. If this is so, then it seems that the

manner in which a state is formed does not place a limit on whether or not we are prepared to classify it as a belief. Thus, (Rw), which in part consists in a claim about the formation of belief, does not seem to be constitutive of belief.

2.7 Transparency and Correctness

So far, then, we have maintained that (R) can capture transparency. Shah (2003) does not consider this kind of counterfactual construal of regulation and hence he concludes that the only remaining option for explaining transparency is the correctness conception of the claim that beliefs aim at truth. Even though it is not possible to motivate the correctness condition along these lines, however, it is nonetheless worth considering whether correctness has a role to play, and we will do so by considering Shah's suggestion.

Now, if we recall, Velleman (2000) originally claims that correctness can be *derived* from regulation. However, Shah (2003) maintains that because Velleman allows that beliefs need not be regulated by the conscious aim of a believer, but can also be regulated unconsciously for truth, Velleman is wrong to conceive of the correctness condition – (C) – as derivable from (Rw). Shah (2003) convinces Velleman (2000) that this is the case and in a joint article on the matter Shah and Velleman (2005) articulate their considered view. They maintain that, if the belief that *p* were always regulated by a literal aim on part of the believer, it would be possible to derive the correctness condition from the fact that the believer aims at truth. However, given that beliefs are often regulated unconsciously, they claim that we cannot derive correctness from regulation, because "how an attitude ought to turn out is not necessarily determined by how it is [weakly] regulated" (Shah and Velleman, 2005, p. 499).[12] Thus, they conclude that we must have a two-part analysis of the concept of belief. They maintain that we ought to think of (Rw) as constitutive of belief to account for those beliefs that are not strongly regulated for truth. As we have seen it is not obvious that this is the case, but they do not see any other way in which such beliefs could be classified as aiming at truth. Furthermore, they think that given that we cannot derive the correctness condition from (Rw), we must think of (C) as an

---

[12] In fact, it is unclear whether correctness could be derived from regulation even on the assumption that the believer possessed a literal aim on every occasion of regulation. As Steglich-Petersen (2006b, p. 500) says, perhaps all that can be derived from the idea that the believer aims at truth is simply that they have "failed to do what… [they] tried to do if the proposition is false, and succeeded in doing what… [they] tried to do if the proposition is true."

additional constitutive feature of belief. They think that although (Rw) alone cannot capture transparency, adding (C) will explain why the question as to *whether to believe that p* immediately collapses into the question as to *whether p* in the context of doxastic deliberation. They claim that this is because one is employing the concept of belief in doxastic deliberation, and insofar as one understands this concept, one recognises that one's belief that *p* will be *correct* if and only if *p*.

Now, a significant criticism of this idea has been voiced by Owens (2003), Steglich-Petersen (2006b), and McHugh (2011c). Shah and Velleman think that in the context of doxastic deliberation, it is the fact that one recognises that it is constitutive of the belief that *p* that it is correct if and only if *p* that explains transparency, because they think that:

"…accepting truth as the standard of correctness for belief involves accepting a directive to believe that *p* only if *p* is true". (Shah, 2003, p. 476, fn 2)

And, that:

"…from a deliberative point of view, making a judgment about what one has reason to believe is bound to have motivational force, because the function of deliberation, as opposed to mere reflection, is to come to a decision or belief on the basis of one's appreciation of reasons." (Shah, 2003, p. 472)

Now, Shah and Velleman's opponents maintain that it is true that on certain meta-normative accounts, making a normative judgement is understood as necessarily having *some* motivational force. However, they claim that it is completely implausible that *accepting a directive* could be understood to have such strong motivational force that it excludes even the possibility of acting otherwise. To see this, we might ask ourselves the following question. Even if the doxastic deliberator accepts that their belief will be correct if and only if true and hence accepts a directive to believe in accordance with truth, what is to stop them thinking as follows?

Well, I know that beliefs are correct if and only if true, but I'm also aware that they can sometimes aim at goals other than truth – like what the believer desires to believe – so why shouldn't I just be a little bit naughty this time? I know I won't have a correct belief, but after all, I'd rather have a soothing one!

Steglich-Petersen (2006b, pp. 506-507) provides an instructive analogy with a moral case:

"It is plausibly regarded as a conceptual truth about promises that promising to ∪ makes it the case that one ought to ∪, at least in the pro tanto sense of 'ought'… This plausibly means that one only counts as promising to ∪ if one acknowledges that one ought subsequently to ∪. But does that mean that one will necessarily go on to ∪? It seems not: one may fail to carry out what one promises to do, otherwise it would not be possible to break a promise. Does it mean, perhaps, that one necessarily intends to ∪ in promising to ∪? This does not seem to be the case either. All it takes for me to count as promising to ∪ is that I acknowledge that I would be subject to negative evaluation, i.e., would have done something wrong and objectionable in not subsequently ∪ing, other things being equal."

So, Shah and Velleman's opponents conclude that simply in virtue of grasping that the belief that *p* is correct if and only if *p*, one need not be motivated to the exclusion of all else to consider solely whether *p* in considering whether to believe that *p*. Hence correctness cannot explain transparency.

However, this line of criticism of the idea that correctness can explain transparency neglects to observe the distinction between being answerable to *epistemic* norms, and being answerable to *practical* norms. We can see this if we consider Steglich-Petersen's analogy with the case of promising. In the moral case, although I may judge that it is correct to ∪, assuming I am rational, whether or not I will then ∪ is subject to practical norms. So, in such a case, I *can* think *I know it would be correct to* 2 *, but I really don't want to*, for example. This is because moral reasoning is a subsection of practical reasoning. However, as we will argue in chapter four, belief is not answerable to practical norms; it is answerable solely to epistemic norms. For this reason, when the doxastic deliberator considers *whether to believe that p* this question amounts to a question about what *epistemic* norms dictate, and in understanding this, it would not be coherent for them to consider anything other than *whether p* because epistemic norms are concerned solely with truth. Whereas practical norms allow for consideration about what one desires, for example, epistemic norms do not. Now, one might object that we have seen that beliefs often *are* governed by practical norms outside of our conscious awareness that this is so:

whatever we might think about the reasons for our beliefs, whether we hold them is often influenced by whether it is pleasant for us to do so etc. However, we must distinguish between being *governed* by a norm and being *answerable* to it. Although beliefs are often illegitimately *governed* by practical norms without our knowing so, in understanding the concept of belief, we understand that beliefs are *answerable* solely to epistemic norms. Now, at this point, one might wonder *why* the doxastic deliberator can only think about the norms to which belief is answerable, rather than also considering the norms by which it can be governed. The answer here lies in understanding the nature of the deliberative standpoint itself. The question *whether to believe that p* does *not* mean *can I believe that p*; it does *not* mean *will I believe that p*: it means *should I believe that p*. Given this normative aspect of the deliberative standpoint, in thinking about what I should believe I can only think about the norms to which belief is answerable: I would no longer be deliberating if I did otherwise. In both the practical and the theoretical case, what it is to deliberate is for one's thinking to be answerable to the relevant norms. But, in understanding that I am deliberating with respect to *belief*, as opposed to a practical matter, I understand that this deliberation is answerable solely to epistemic norms. The position of the doxastic deliberator is not like that of someone who asks *should I treat my friend fairly or unfairly*. It is rather like the standpoint of someone who, having already resolved to treat their friend fairly, asks themselves *what should I do*, since such a person has made it the case that like the doxastic deliberator, their deliberation is only framed by one set of norms (in this case, moral rather than epistemic). In the same way that such a person's deliberation can only be responsive to considerations as to what would count as fair treatment, so the doxastic deliberator's thinking can only be responsive to considerations pertinent to *whether p*. The difference between the moral and the doxastic case is simply that the doxastic deliberator, given the nature of epistemic norms, automatically occupies this standpoint, whereas the moral deliberator, given the nature of practical norms, must first resolve to have their thought and action guided solely by moral norms in order their deliberation to acquire transparency.

So, it seems that Shah and Velleman (2005) are right that transparency can be explained by appeal to the idea that in understanding the concept of belief they are employing, the doxastic deliberator understands that their belief that *p* will be correct

if and only if *p* is true. At this point, one might wonder whether we have really established that the correctness condition is *constitutive* of belief. Might it not be an extrinsic feature of belief that it is answerable to epistemic norms? Gideon Rosen (2002, p. 617) effectively presses this point:

"The indisputable fact that belief is… [answerable to] norms is consistent with the following picture: The doxastic facts are constituted entirely from non-normative materials. But once in place, they engage with an independent body of cognitive norms. On this view, if God were to suspend the norms of (say) rationality, he would not annihilate belief. He would simply suspend some of the constraints that govern it."

The question is whether the fact that the belief that *p* is correct if and only if *p* is a fact about belief itself, or a fact about the norms that pertain to belief, and its relation to those norms. If it is a fact about the norms, then, as in the above picture, it would be possible to remove those norms – i.e. the fact that the belief that *p* is correct if and only if *p* – without thereby getting rid of beliefs. However, it is highly counterintuitive that the states that would survive this normative extinction would count as beliefs. Such states would be entirely impervious to rational criticism: it would not be possible to say of someone e.g. *he ought not to believe that p*, but this clearly breaches our folk psychological intuitions about what counts as a belief. So, it seems that not only are beliefs answerable to epistemic norms, but they are constitutively so.

But haven't we already explained transparency solely in terms of (R): It is constitutive of belief that if it is consciously regulated, it is so-regulated solely for truth? Yes, and given this there would be no need to *add* the correctness condition as Shah and Velleman (2005) maintain. However, it seems that if (R) were proffered as a complete explanation of transparency, one might be left with a sense of dissatisfaction. One might be tempted to ask: but *why* does the doxastic deliberator conceive of belief as necessarily regulated in this way? What is it about belief that explains why it must be so-regulated? It is now time to propose the answer to this question in terms of the idea that belief is at base a normative phenomenon. As McDowell (1994, p. 60) maintains:

"…even in the case of a belief that one simply finds oneself with, the question of one's entitlement to it can always be raised."

And, no matter how one arrived at one's belief, the question of one's rational entitlement to one's belief can always be raised because:

(C): It is constitutive of belief that it is correct if and only if its contents are true.

We have seen Velleman (2000) maintain that correctness can be derived from regulation, and Shah (2003) deny this, leading Shah and Velleman (2005) to affirm that both (C) and (Rw) are independently constitutive of belief. However, we have now arrived at a conception of belief according to which our regulation claim – (R) – is derived from (C).

In sum then, we have argued that the fundamental explanation of transparency is – as Shah and Velleman (2005) claim – that in understanding the concept of belief, the doxastic deliberator understands that their belief that p will be *correct* if and only if p. However, *contra* Shah and Velleman (2005), we have argued that it does not seem that the regulation condition is a separate constitutive condition of belief. Rather, we can derive our regulation claim, which the doxastic deliberator grasps in the context of doxastic deliberation, from (C).

2.8 Back to Juliet and the Precipice

Before we close this chapter, let us return to the initial problem cases with which we began in chapter one and examine how what we have established here bears upon our assessment of those cases. In chapter one, we saw that maintaining that belief and reasons are somehow importantly interrelated enables us to account for the patterns in the behaviour of both Juliet and the woman in Precipice. All of their avowals and considered behaviour hang together, as does all of their unconsidered behaviour. We set out to explore the idea that this is because their avowals and considered behaviour pertain to what they believe, their unconsidered behaviour to other mental states, and we are now in a position to say more about the particular relevance of their considered behaviour to what they believe.

In this chapter, we have argued that belief is a fundamentally normative concept: it is constitutive of believing that p that one's belief is correct if and only if p. We have

also argued that this explains the fact that it is constitutive of belief that if beliefs are consciously regulated for truth, they are so-regulated solely in accordance with truth.

Now, it is important to note that such regulation need not be covert: it need not go on only in the privacy of a person's mind. It can also take the form of an overt rational defence. And, if we turn to Juliet's case to begin with, we see that *ex hypothesi*:

"She is prepared to argue coherently, sincerely, and vehemently for equality of intelligence and has argued the point repeatedly in the past." (Schwitzgebel, 2010, p. 532)

Now, what is this behaviour other than the conscious regulation of her belief that the races are intellectually equal for truth? Similarly, in Precipice, although the woman "didn't dare move toward the transparent center of the walkway", she nonetheless *remains* on the walkway (Gendler, 2008b, p. 634). Such considered behaviour can also be naturally construed as the woman keeping her beliefs regarding her personal safety in line with her epistemic reasons – that is, consciously regulating her belief for truth.

Now, it is important to note that this is *not* to say that simply because someone asserts or argues that p, or ensures that they behave as if p is the case, they thereby believe that p. This is evidently not the case: consider the behaviour of a liar, for example. They assert that p and behave in other considered ways as if p but do not believe that p. However, it would be difficult to construe the case of the woman in Precipice as one in which she is insincere in keeping herself up on the walkway. And it is part of the description of Juliet's case that she is sincere in her assertions.

Furthermore, our assessments of Implicit Racist and Precipice are not meant to imply that Schwitzgebel is incorrect to think that one's unthinking behaviour is often revelatory of what one believes. Under normal circumstances, when there are no excusing habits or fears in play, for example, one's unthinking behaviour will often be in line with what one believes. That is, it will be revelatory of the fact that one possesses the disposition to consciously regulate one's belief in accordance with one's epistemic reasons. Imagine a case of an unconscious perceptual belief, for example, such as the belief that one's wine glass is in front of one. Now, the only evidence we might have that one believes this is one's reaching out to take a sip from

the glass. Such unthinking behaviour is relevant to the question of what a subject believes insofar as it evidences how they would consciously regulate their belief if it were drawn to their attention – if necessary, they could provide reason(s) in its defence such as *I see it*, *I have no reason to doubt my senses* etc. And, it is this disposition (ultimately explicable in terms of belief's correctness condition) that constitutes their belief. All we are claiming is that in these kinds of cases, such unthinking behaviour is itself merely revelatory and not constitutive of belief in the way in which conscious regulation or the disposition to it is.

A further important point to note is that what we are saying is perfectly compatible with the idea that under certain conditions, one's unthinking behaviour is *more* pertinent to what one believes than one's considered behaviour. Herein lies the truth of the expression *actions speak louder than words*. Consider a case in which someone is insincere in their assertions and is caught out by their unthinking behaviour, for example. It seems that the pull towards ascribing Juliet the belief that it is not the case that the races are intellectually equal comes when one conceives of her case as one involving such insincerity. Although this is not Juliet's case *ex hypothesi*, this is not to deny that this kind of case exists. In such cases, when we say of someone that their actions speak louder than their words, we are denigrating their words to the status of *mere* words – that is to say, we refuse their words the status of avowal of belief. Part of the wisdom behind the expression is that verbal behaviour has no automatic purchase on belief, as we have seen. As for the actions themselves, they acquire the very status that the person's verbal behaviour is denied: they speak. That is, they reveal what the person takes their reasons to be (although they nonetheless do not *constitute* the person's taking their reasons to be that way, as considered behaviour often does).

Cases in which one's words are mere words are in fact the converse of a case such as Juliet's. We can think of Implicit Racist and Precipice as on one side of a continuum as to which kinds of behaviour are pertinent to what a person believes. Such cases involve excusing conditions that mitigate the relevance of a person's unthinking behaviour to what they believe. In the middle of the continuum we have cases in which there are no excusing conditions in play and hence both a person's considered and unconsidered behaviour is pertinent to what they believe – although the former is constitutive, the latter only revelatory of what is constitutive. Finally, at the other end

of the spectrum are cases in which there are excusing conditions in play – such as the person's insincerity – that undermine the relevance of their considered behaviour to our proper assessment as to what they believe.

We have acknowledged that in the middle of the continuum, where there are no excusing conditions, all of a person's behaviour will hang together, signalling what they believe. Nonetheless, such cases are rare, because human beings are complicated creatures, with many different mental states interfering with one another at any one time.

Conclusion

In this chapter we set out to explore the relationship between belief and reasons, and to provide some support for our claim in chapter one that it is a significant one for understanding the nature of belief. We noted that the phenomenon Shah (2003) refers to as the transparency of doxastic deliberation seems to support the idea that reasons are important for belief. Transparency is the idea that the phenomenology of consciously considering *whether to believe that p* is such that this question is immediately – non-inferentially – transparent to the question *whether p*. However, we also observed another phenomenon (which we dubbed *heterogeneity*) that seems to suggest that epistemic reasons do *not* stand in a privileged position with respect to belief. This is the fact that in spite of transparency, outside of conscious awareness, our beliefs can be regulated by a number of factors other than truth – such as what we desire to be the case, for example. This phenomenon seems to support Schwitzgebel's contention that there is no significant relationship between belief and reasons and hence that all of our behaviour – considered and unconsidered – is on a par when it comes to evaluating whether we hold a certain belief (what we have called his egalitarianism). So, we were presented with a task: to explain transparency in the face of heterogeneity in a way that would reveal that there is nonetheless a significant relationship between belief and reasons. We set about doing so by considering Williams' (1973) famous and suggestive claim that *beliefs aim at truth*.

The broad idea we were exploring is that it is because the doxastic deliberator understands that beliefs aim at truth that when considering *whether to believe that p* this question is transparent for them to the question *whether p*. We outlined the two main interpretations of the claim that beliefs aim at truth as follows: first, as a claim

61

about the manner in which belief is *regulated*, secondly, as a claim about belief's correctness condition – that the belief that *p* is correct if and only if *p*. We saw that Shah (2003) argues that we cannot account for transparency by appealing to the manner in which beliefs are regulated because, according to him, it cannot be said to be constitutive of belief that it is strongly regulated for truth. Given that in understanding the concept of belief the doxastic deliberator would know this, regulation seems unable to explain transparency. Why couldn't the doxastic deliberator understand that beliefs are often regulated for factors other than truth and hence consciously consider such factors in deliberating about *whether to believe that p*? We then turned to Steglich-Petersen's (2006b) attempt to defend the idea that we *can* capture transparency (in the face of heterogeneity) solely by appeal to how beliefs need be regulated. He attempts to do so by maintaining there to be *two* main concepts of belief: what he calls that core concept of belief, according to which belief is strongly regulated for truth, and the empirical concept of belief, according to which belief is only weakly regulated for truth. Steglich-Petersen (2006b) maintains that it is because the doxastic deliberator employs the core concept of belief that their deliberation exhibits transparency. However, we objected that this explanation is not satisfactory. The problem of accounting for transparency in the face of heterogeneity simply re-emerges for Steglich-Petersen (2006b) when we ask him *why* the doxastic deliberator employs the core concept of belief, as opposed to the empirical one. In order to avoid this sort of objection, we proposed an alternative explanation in terms of regulation, with a different focus: rather than focussing on how all beliefs are regulated both consciously and unconsciously, our account is concerned with how all beliefs *would* be consciously regulated. We argued that although some beliefs will never be consciously regulated, if and when they are, in order to count as beliefs, they need be so-regulated solely for truth. Thus, we contested Shah and Velleman's (2005) claim that "no single interpretation of 'truth-regulation' can explain both… [heterogeneity and transparency]". Our final observation with respect to regulation is that it is unclear that we need to add that (Rw) is constitutive of belief. We considered a case in which an evil demon causes someone to form a belief, whose contents they nonetheless currently consider true. We noted that in such a case, it seems that we would not wish to withhold the status of belief from this mental state.

We then went on to consider whether correctness has a role to play in explaining transparency. We saw that Shah (2003) convinces Velleman (2000) that this is so and that Shah and Velleman (2005) maintain that it is because, in understanding the concept of belief, the doxastic deliberator understands that their belief will be *correct* if and only if true, that the question *whether to believe that p* is phenomenologically transparent for them to the question *whether p*. However, we then saw that many (Owens (2003), Steglich-Petersen (2006b), McHugh (2011c)) have objected that transparency cannot be accounted for in terms of a correctness condition in this way, because it is prohibitively implausible that the comprehension of such a correctness condition would necessitate the doxastic deliberator's being motivated to comply with it to the exclusion of all other considerations. However, we argued that such a response to Shah and Velleman emerges from a failure to appreciate the difference between the manner in which actions are answerable to practical norms, and beliefs are answerable to epistemic norms. We maintained that considering *whether to believe that p*, simply is considering the question *should I believe that p*. Once we recognise this, and we understand that belief is answerable solely to epistemic norms, it becomes clear why doxastic deliberation is transparent: one is asking oneself *should I believe that p* and one recognises that all that is relevant to this question is *whether p*. So, we arrived at a picture of transparency as ultimately explicable in terms of the idea that it is constitutive of the belief that *p* that it is correct if and only if *p*. This constitutive feature of belief in turn explains the previously proposed constitutive feature of belief, namely, that if belief is consciously regulated for truth, it is so-regulated solely for truth.

We closed by exploring the implications of where we have arrived in this chapter for our previous evaluations of Implicit Racist and Precipice. We can now defend the idea that Juliet and the woman in Precipice hold the relevant beliefs on the basis that their considered behaviour is not only revelatory of their holding such beliefs, it is constitutive of it. This behaviour simply *is* their conscious regulation of their beliefs, or at least reveals that they are disposed to consciously regulate the state in question in accordance with their subjective epistemic reasons, and hence that it is a belief.

Having arrived at a conception of belief according to which it is both constitutive that it be regulated in a certain way under certain conditions, and that it is correct if and only if its contents are true, it is now time for us to examine exactly what these

two conditions involve in greater depth. We will begin in the next chapter, chapter three, by exploring the regulation condition further, and then, in chapter four, we will turn to the correctness condition.

3. The Regulation Condition

In the previous chapter, in search of a defence of the idea that there is a significant relationship between belief and reasons, we looked to the transparency of doxastic deliberation, which seemed to represent good evidence that there is such a relationship. We concluded that Williams' (1973) famous aphorism that *beliefs aim at truth* can be construed so as to explain transparency in the face of heterogeneity. We argued that this aphorism is best understood at the most fundamental level as a claim about belief's correctness condition, but that from this correctness condition we can derive a further constitutive feature of belief – about belief's regulation. We called this feature (R) and it is the idea that if a belief is consciously regulated for truth, it is so-regulated solely for truth. In this chapter, we will investigate precisely how (R) is to be understood. In particular, we will consider whether it should be understood as involving the *believer* possessing a doxastic aim. All parties to the debate so-far have assumed that conscious regulation for truth would involve such an aim on the part of the believer if belief itself is to be considered as aiming at truth. However, we will argue that the believer themselves possesses no such aim but that their belief nonetheless aims at truth in accordance with the manner in which it is consciously regulated. We will also see that, as Williams (1978) maintains, there is no distinction between conceiving of belief as aiming at *truth* and conceiving of it as aiming at *knowledge* in this sense.

We will proceed by examining two further features of doxastic deliberation.[13] The first of these features is what Steglich-Petersen (2009) has called *exclusivity*, and the second, what McHugh (2011a, 2011c, 2012b) has called *demandingness*. Roughly speaking, exclusivity is the idea that when one considers *whether to believe that p*, one can only think of oneself as moved to form the belief that *p* on the basis of epistemic reasons. Demandingness, on the other hand, is the idea that when one considers *whether to believe that p*, one can only think of oneself as moved to form the belief that *p* on the basis of *sufficient* epistemic reason that *p*. It is important to note that – like transparency – neither of these features of doxastic deliberation

---

[13] These features, like transparency, seem to pertain to doxastic deliberation framed by the question *whether to believe that p* not solely when one has not previously believed that *p*, but also when one is considering whether to continue to, or cease to believe that *p*: that is, all forms of conscious belief regulation. Nonetheless, for simplicity's sake, we will assume that the deliberator is in the former circumstance.

conflict with heterogeneity. It can be that one is being influenced (unbeknownst to oneself) by non-epistemic reasons, and that these reasons can lead one to think of oneself as having sufficient epistemic reason.

We will begin by outlining the differences between transparency, exclusivity, and demandingness. We will then consider Owens' (2003) influential argument that given that doxastic deliberation exhibits exclusivity, the deliberator themselves cannot be considered to be aiming at truth; in fact, they cannot be considered to have an aim *of any kind*. Owens argues that this is because their purported *aim* would not *integrate* with their other aims in the manner of ordinary, non-epistemic aims. We will then examine a line of response to Owens' argument, which involves denying that we need think of doxastic deliberation as exhibiting exclusivity. We will argue that this line of response is not convincing. Given this, we will explore an alternative manner of conceiving of the doxastic deliberator's aim – on the model of their aiming to ∪ *iff* they are satisfied that *p*. We will see that this kind of aim provides us with an example of deliberation concerning a non-epistemic aim that exhibits an analogue of exclusivity. However, we will argue that we cannot model the doxastic deliberator's aim in this way because it would be rendered circular: to consciously believe that *p* simply *is* to be consciously satisfied that *p*. And, given that being consciously satisfied that *p* is taking oneself to *know* that *p* (as Owens (2000) and McHugh (2011c) have urged), thinking of the doxastic deliberator as thereby aiming at knowledge is also precluded. So, we will conclude that the doxastic deliberator themselves does not possess an aim. Rather, their belief itself aims at truth insofar as what it *is* to form this belief is to become satisfied that *p* is true. Furthermore, as Williams (1978) maintains, there is no distinction between maintaining this to be the case and holding that beliefs themselves aim at knowledge: being satisfied that *p* is taking oneself to know that *p*.

3.1 Transparency, Exclusivity, and Demandingness

Let us begin by sharpening our understanding of the distinctions between transparency, exclusivity, and demandingness, as having a firm grasp of these terms will be important in what follows. To remind ourselves, transparency is the idea that when considering *whether to believe that p*, this question immediately (non-inferentially) gives way to the question *whether p*. Now, as for exclusivity and

demandingness, things are a little more complicated than we intimated in the introduction insofar as we can construct both a *descriptive* and a *normative* version of both of these claims. The descriptive version of exclusivity is the idea that the doxastic deliberator *can* only think of themselves as moved to believe that *p* solely on the basis of epistemic reasons. The normative construal, on the other hand, has it that the doxastic deliberator *should* base their belief exclusively on such considerations. The descriptive version of demandingness is the idea that the doxastic deliberator can only think of themselves as moved to believe that *p* on the basis of *sufficient* epistemic reason in favour of *p*; the normative claim is that they *should* only form the believe that *p* on the basis of sufficient epistemic reason. In this chapter, we will be interested in the descriptive versions of both of these claims – we will turn to the normative versions in the following chapter.

Now, the distinction between demandingness and transparency is relatively clear: transparency is a claim about what one is bound to consider when one considers *whether to believe that p*; demandingness, on the other hand, is a claim about the *amount* of epistemic reason one must have in favour of *p* before one can take oneself to form a belief that *p* on this basis. However, the difference between exclusivity and transparency is a subtle one and may be less obvious. It seems that transparency – the claim that all the doxastic deliberator can consider are epistemic reasons – *entails* exclusivity – the claim that they can only think of themselves as forming their belief on the basis of epistemic reasons. However, transparency is a *stronger* claim than exclusivity: transparency says that one cannot even consider non-epistemic reasons as relevant to the question as to whether to believe that p, whereas exclusivity has it that one cannot think of oneself as going ahead and forming a belief on this basis. So, exclusivity is consistent with the idea that one could consider non-epistemic reasons as relevant to *whether to believe that p* – according to exclusivity one simply cannot knowingly get oneself to form a belief on this basis.

3.2 Believing and Guessing

Having strengthened our hold on the distinctions between these three phenomena, we will now move on to consider Owens' (2003) argument that the exclusivity of doxastic deliberation renders it inappropriate to think of the doxastic deliberator as possessing an aim. Owens mounts his argument by contrasting deliberation about

whether to believe that p with deliberation about *whether to guess that p*. He argues that whilst the guesser ought to be considered to be aiming at truth, the doxastic deliberator ought not.

He begins his discussion by pointing out that what one can guess is clearly *constrained* by what one takes to be true. For example, it is evidently impossible for me to guess that I am 105 years old, knowing as I do that this is false. He claims that:

"Clearly, truth is the standard of correctness for a guess and, I maintain, what explains this is the fact that a guesser intends to guess truly. The aim of a guess is to get it right: a successful guess is a true guess and a false guess is a failure as a guess. Someone who does not intend to guess truly is not really guessing." (Owens, 2003, p. 290)

Nonetheless, what one can guess can be consciously influenced by factors other than truth, as Owens illustrates. He asks us to imagine being in a quiz situation in which we are invited to guess whether the Earth's population is greater than seven billion and will win $1 million for the correct answer. In such a case, it is clear that even if one has very little epistemic reason in favour of the claim that the Earth's population is greater than seven billion, one can make a guess that it is because one desires to win the money. In this situation, it is clear that the guesser weighs their truth aim *qua* guesser to answer the question correctly against their other aims – most prominently their aim to win the money. So, their guess is based upon *both* epistemic and non-epistemic reasons.

However, Owens affirms that *belief* does not work like this. Although one can straightforwardly *guess* that the Earth's population is greater than seven billion in spite of little epistemic reason one way or the other, it would be *impossible* for one to form a belief under such circumstances. Owens thinks that this is because one would be unable to conceive of oneself as basing a belief about the Earth's population in part upon non-epistemic considerations such as the value of winning $1 million as one can do with a guess. However much one wants the $1 million, one *cannot* go ahead and form a belief about the Earth's population under such circumstances. This is not to say that Owens denies heterogeneity. He fully acknowledges that outside of conscious awareness, non-epistemic factors can and do influence belief formation. He simply claims that such factors cannot be the basis upon which we consciously

form beliefs. That is, doxastic deliberation exhibits exclusivity, whereas deliberating about *whether to guess that p* does not. According to Owens, one can integrate epistemic and non-epistemic considerations when it comes to guessing, but not when it comes to believing.

3.3 Our Ordinary Conception of an Aim

Now, Owens thinks that it is precisely because the guesser's truth aim integrates with their other (non-epistemic) aims in the manner described above that it is apt to think of guessing as aiming at truth. Likewise, in the case of belief, our *inability* to let any concern other than truth explicitly move us to believe renders it inappropriate to consider the doxastic deliberator as aiming at truth. As he puts it:

"If this talk of purpose is to be more than an empty metaphor, the notion of purpose invoked must be one that does explanatory work for us outside the domain of epistemic norms. And, surely, it is our ordinary notion of a purpose which the truth-aim theorist means to be employing." (Owens, 2003, p. 295)

So, Owens does not explicitly claim that the kind of integration he has in mind is *essential* to or *necessary* for something's being an aim, but in effect he proceeds as if this were the case, satisfied that our *ordinary* notion of an aim or purpose (again, these terms will be used interchangeably) includes this idea.

To illustrate the ordinary kind of integration of aims Owens has in mind, let us consider an example. Imagine that Paul aims to learn French. Perhaps the *best* way for Paul to achieve this aim is to move to France, and spend every waking hour there either conversing in French or pouring over French grammar books. Nonetheless, if Paul were to move to France, this would conflict with his other aims – holding down his job in England, keeping his children in their English schools etc. Accordingly, he strikes a compromise and enrols himself in a French evening class at his local university. Now, the fact that Paul does not drop everything and move to France does not mean that he does not *really* aim to learn French – despite the fact that he is aware that moving to France would be the best way to realise this aim. It simply means that in so far as Paul is rational, his intention to learn French must take its place in the hierarchy of his other aims, and be integrated accordingly.

Now, Owens' idea is of course that this is how we should expect *all* instances of aiming to behave. And, it seems that the doxastic deliberator's purported aim does *not* work in this way: doxastic deliberation exhibits exclusivity. Given that this is so, we cannot claim to be shedding light upon doxastic deliberation and belief by appealing to the similarity between the doxastic deliberator's *aim* and non-epistemic aims. This is how Owens puts pressure on the very idea that the doxastic deliberator has an aim.

3.4 To Deny Exclusivity?

The obvious solution to this threat to the idea that the doxastic deliberator is genuinely aiming is to deny the assumption upon which it is based – that doxastic deliberation exhibits exclusivity (which would also involve a denial of the stronger claim that doxastic deliberation is transparent), a strategy recently adopted by McHugh (2012b). However, we will now argue that this option has not been rendered plausible.

Let us begin by re-considering the case Owens employs in order to illustrate the exclusivity of doxastic deliberation. He asks us to imagine being offered $1 million as a reward for forming a belief about the Earth's population for which we have very little evidence. His thought is that we cannot do so as we cannot consciously base a belief on a non-epistemic reason such as our desire for the money. However, these kinds of example do not establish that the doxastic deliberator can only consciously regulate their beliefs on the basis of epistemic reasons because, as McHugh (2012b, p. 8) puts it:

"The fact that you cannot win the reward in such examples does not provide evidence for exclusivity, because that fact is already predicted by demandingness. In order to see whether exclusivity is genuine, we must look at cases where a subject is considering whether to believe a proposition for which the evidence is, by her lights, sufficient. It is only in these cases that the symptoms of demandingness and of exclusivity could come clearly apart. Exclusivity predicts that, even in such cases, non-evidential considerations cannot be motivating reasons for belief (weak exclusivity), or for either belief or withholding (strong exclusivity). Demandingness makes no such prediction."

As we can see, McHugh also draws a subdivision in the category of exclusivity between what he dubs *weak* and *strong* versions of it. According to weak exclusivity, non-evidential reasons cannot be motivating reasons for belief. According to the stronger version of the claim, non-evidential reasons cannot be motivating reasons for either belief or for the withholding of belief. Now, the thought is that cases such as Owens' $1 million case cannot establish either weak or strong exclusivity, because they are cases in which the condition of demandingness is not met. We cannot tell whether the person is unable to believe or withhold belief because they cannot consciously do so on the basis of non-epistemic reasons, *or* because they cannot consciously do so whilst being aware that they have insufficient epistemic reason. Keith Frankish (2007, p. 540) puts the point in terms of *enabling conditions*. That is, for all examples like Owens' show, it may well be that the doxastic deliberator can base their belief or withholding of belief in part upon non-epistemic reasons, but that they can only do so when certain enabling conditions are in place: namely, when they take themselves to have sufficient epistemic reason to form the relevant belief. So, in order to consider whether the doxastic deliberator can base their belief or withholding of belief in part upon non-epistemic reasons, we had better consider cases in which they think that they have sufficient epistemic reason to form the belief.

Before we do so though, let us note that it does seem that demandingness is true of doxastic deliberation. Imagine a case in which, having consulted all of the available meteorological evidence, it seems to you that there is a .6 chance that it will rain today and a .4 chance that it will remain dry (McHugh, 2011c). In such a case, it would be impossible to go ahead and believe that it will rain today. One may well be able to form the belief *there is a .6 chance that it will rain today*, and even adopt a .6 *credence* that it will rain today, but one will find oneself unable to form the outright belief *it will rain today* on the basis of this evidence. So, whatever we say about the types of reasons that can knowingly move one to believe, it seems that demandingness is a constraint upon doxastic deliberation: one must consider oneself to have sufficient epistemic reason in favour of p to be able to form the belief that p as the result of doxastic deliberation. Believing is certainly distinct from guessing in this respect: one need not think of oneself as having sufficient epistemic reason in

favour of *p* to guess that *p* – indeed, thinking of oneself as having such evidence may well *preclude* guessing that *p*.

But what, exactly, counts as *sufficient* evidence? Philip Nickel (2010, p. 315) draws up a scale of different grades of evidence, according to which:

"At the top of the scale are those claims infallibly known to be true, either because (AI) their truth is made evident by the very fact of their being conceived, or (A2) conceiving them gives a person evidence to believe them that could never be overturned."

Whether or not such forms of evidence indeed exist, this is not the kind of evidence we have in mind when we talk about sufficient evidence. Rather, we are conceiving of sufficiency as something like the following:

"…more than having better or stronger evidence for the proposition than for its negation… some high degree or strength of evidence, or some particular kind of evidence, for the proposition." (McHugh, 2012b, p. 6)

Now, it is important to note that in order to cast doubt upon either weak or strong exclusivity, it is necessary to accept the following idea:

"…there is a distinction between reasons that are sufficient to permit belief and reasons that are sufficient to compel it." (Frankish, 2007, p. 541)

That is, in some (but not all) cases of belief formation, one's evidence for *p* can be such that it is sufficient to *enable* belief in *p* but does not *necessitate* it. A belief formed on the basis of a perception in a typical scenario provides a clear example of a case in which this kind of *discretion* (as McHugh (2012b) calls it) does not seem to apply. I see the wine glass in front of my eyes clear as day; withholding belief as to whether or not it is there is not a real possibility for me. Nonetheless, it has seemed to some that there are cases in which discretion holds. Philip Nickel (2010, p. 315)[14] provides the following example, loosely based on one of Austin's (1961, p. 51):

"…a person identifies a bird he sees relatively clearly as an American goldfinch, but he has some reason to think that another similar bird might have flown into the area, contrary to its usual habits (due to global warming, say) and the birdwatcher has only

---

[14] Although Nickel (2010) affirms discretion, he does not deny exclusivity.

weak evidence to believe that he could distinguish this other bird from the American goldfinch."

The idea is that in such a case, the person has sufficient evidence to think that he sees an American goldfinch, but his belief is not compelled by his evidence.

Now, it seems that the only way to evaluate both discretion and the denial of weak and/or strong exclusivity is to consider some further cases. Let us begin with some examples that seem to cast doubt upon strong exclusivity, i.e. the claim that one cannot even withhold belief if one takes oneself to be doing so on the basis of non-epistemic reason. The first of these pertains to a case involving friendship:

"**A.** Your friend is accused of some terrible wrong, and several seemingly independent witnesses have reported that he did it. By your lights, you would not easily have this evidence were your friend not guilty. It would require that all of the witnesses be either mistaken or lying—a possibility that cannot strictly be ruled out, but is far-fetched. However, you value the friendship and you think that, if you were to believe in your friend's guilt, it would have profoundly damaging effects on the friendship, particularly should your friend turn out to be innocent after all."
(McHugh, 2012b, p. 12)

In such a case, one might think that although it is not open to you to believe that your friend is not guilty, you are nonetheless able to take into account a non-evidential consideration – "the potential damage of believing that your friend is guilty" – and withhold belief as to whether your friend is guilty or innocent "remaining open to the possibility that the witnesses are mistaken or lying, for this reason" (McHugh, 2012b, p. 12).

Cases involving friendship are complicated, however. It seems that our intuitions regarding whether or not we can knowingly form a belief on the basis of the non-evidential consideration regarding "the potential damage of believing that your friend is guilty" may well be muddied by the fact that we will presumably be inclined to take our friendship *itself* as evidence in favour of our friend's innocence. That is, we will presumably include evidential considerations such as *A friend of mine wouldn't do something like that!* in our deliberations, as well as our specific knowledge of our friend's character: this is the role of character witnesses in a court of law, of course.

Furthermore, because of the emotional attachment involved by definition in cases of friendship, one is likely to refuse or be unable (in some sense) to continue to deliberate regarding the guilt of one's friend. This withdrawal from deliberation should be carefully distinguished from deliberating and then withholding belief on the basis of this deliberation. McHugh's (2012b, p.13) second case provides an example free from such complications, so let us focus our attention upon this instead:

"**B.** You have a cheque that you badly need to lodge by the end of the week. It would be much easier to lodge it tomorrow (Saturday) than today (Friday). You have been to the bank on Saturdays before, and thus have what you regard as sufficient inductive grounds to believe that the bank will be open tomorrow. However, it is not impossible that the bank has changed its hours."[15]

Again, one opposing strong exclusivity would claim that in such a case, although ordinarily you would believe that the bank will be open, because it is so important to you that it *is* in this case, you can explicitly ground your withholding of belief upon this non-evidential consideration. But is this example coherent as described? We will now argue that this is not obviously the case. We will do so by diagnosing the intuitive appeal of the idea that such a case is possible in terms of the intuitive appeal of *contextualism* about considering one's reasons sufficient. We will argue that it is plausible that we *think* cases like B. coherent because we are convinced by the plausibility of such contextualism, as opposed to the cases being genuinely coherent as described. And, hence, it has not been established that strong exclusivity is really in doubt. Neither, for that matter, has the discretion assumption been shown to be intuitively plausible.

McHugh (2012b, p. 13) states:

"I am not suggesting that you are always guaranteed to succeed when you try to withhold a belief in this way. You might find yourself stubbornly convinced of the truth of *p*, despite wishing to withhold belief and countenance error possibilities. My claim is only that you can sometimes withhold beliefs in the way described. Non-evidential considerations are not excluded from being motivating reasons for withholding belief."

---

[15] This case is an adaptation of Keith De Rose's (1992) well-known bank case.

This admission on McHugh's part is telling. We might ask *why* reflection upon the importance of the bank's being open will *guarantee* that I withhold belief. It seems that the answer to this question is that we intuitively acknowledge that the only kind of influence explicit considerations of non-epistemic reasons (such as the importance of the bank's being open) can have over whether one withholds belief is *indirect*: it seems that we rely upon the idea that they do so *via* evidential considerations. That is to say, in A., "the potential damage of believing that your friend is guilty" enlivens your mind to the possibility that "the witnesses are mistaken or lying". Similarly, in B., the fact that your cheque badly needs to be lodged by the end of the week makes vivid the fact that "it is not impossible that the bank has changed its hours." (McHugh, 2012b, pp. 12-13)

Indeed, (and this is where De Rose, the original exponent of cases such as B., would approve of our verdict) it seems that the role we think of these non-evidential considerations as playing is as rendering the deliberator's evidence *insufficient* from their point of view. That is, it seems that when we consider cases such as B., we acknowledge that whether one takes certain evidence to be sufficient depends upon the *context*, and although one can be explicitly aware that in another – less demanding – context one would understand the evidence one currently has to be sufficient, in this context, one cannot conceive of it this way as one's mind is enlivened to the reasons for doubting one's evidence. If this is what we are really imagining when we imagine cases such as B., then, of course, it is demandingness that explains why we withhold belief, not discretion and the falsity of strong exclusivity.

But in each case, haven't we *stipulated* that the person in question takes themselves to have sufficient evidence? Yes, but it seems that it is this stipulation that makes McHugh's exact description of the case far from obviously coherent. The charge against those such as McHugh who rely upon examples like A. and B. is that they fire up our *contextualist* intuitions about when one will consider oneself to have sufficient epistemic reason, and then they illegitimately rely upon these intuitive verdicts to reach the conclusion that one can consider one's evidence sufficient and yet withhold belief for non-epistemic reasons.

Now, McHugh (2012b, p. 15) in fact considers this kind of response to his position and tries to discredit it by observing that if we are correct:

"…a practical consideration can influence you to withhold belief, or to lose a belief you previously held, without being a reason for which you do so, and even though there is no change in your evidence."

He then moves from this observation to the claim that:

"**You cannot deliberatively acknowledge the role of the practical consideration in such events of doxastic rearrangement, and yet there is no change in your evidence that could explain them either. It thus becomes puzzling what would be going on from your own point of view, in such a case.**" (McHugh, 2012b, p. 15)

However, although we are claiming that one cannot acknowledge non-epistemic reasons as reasons upon which one bases one's belief, we have acknowledged that they can feature explicitly in one's thinking about one's situation. In a case like B., it seems intuitively plausible that one would understand one's behaviour as follows:

I know that I am only doubtful that the bank will be open because of the importance of its being open on this occasion: I recognise that I have what I would ordinarily – in less demanding contexts – think of as sufficient evidence that this is the case. But nonetheless, the bank might not be open; it might have changed its hours.

This is what is going on from your point of view. And, as soon as we understand it in terms of your flipping between a third- and first-person perspective upon your own situation in this way, it is perfectly intelligible.

Now, as far as weak exclusivity goes, the same considerations hold *a fortiori*. McHugh (2012b, pp. 16-17) presents the following three cases in order to try to put pressure on the idea that weak exclusivity holds:

"**C.** Your friend is accused of some terrible wrong, but he has an alibi supported by several seemingly independent witnesses. By your lights, you would not easily have this evidence were your friend not innocent. It would require that all of the witnesses be either mistaken or lying—a possibility that cannot strictly be ruled out, but is far-fetched. What's more, you value the friendship and you think that, if you did not

believe in your friend's innocence, it would have profoundly damaging effects on the friendship, should he indeed turn out to be innocent.

D. You are playing tennis. You are much stronger than your opponent, so you regard the possibility of defeat as far-fetched. If you believed that you would win, your confidence would make victory even more likely.

E. You are a car salesman. You are trying to sell a particular car. All the evidence suggests that it is reliable—the model is said to be very reliable, and there is nothing to suggest that this particular vehicle is any exception, but you can't absolutely rule out that there is some fault somewhere. You will be more convincing in your sales pitch if you believe that the car is reliable."

In these cases, unlike in A. and B., the idea is that the non-epistemic reasons can explicitly be taken by the person in question as reasons for their belief that *p*, as opposed to their simply withholding belief regarding *p*. However, as with cases A. and B., it seems that our intuitions about the contextualism that is true of considering oneself to have sufficient reason may well be doing all of the work.

In sum then, we have argued that the kinds of cases we have considered do not render discretion plausible or strong and/or weak exclusivity implausible. This is because we can offer an explanation of the intuitive pull of such cases in terms of the intuitive pull of contextualism about regarding oneself as having sufficient epistemic reason. Of course, this is not a knock-down argument that exclusivity is true of doxastic deliberation. Rather, we have simply cast doubt upon the idea that the denial of exclusivity has been rendered plausible by appeal to the kinds of cases we have been considering. We will now move on to consider an alternative strategy to denying exclusivity by which to overcome the difficulty Owens has raised for the idea that the doxastic deliberator possesses an aim.

3.5 A Non-Epistemic Analogy

As we have seen, Owens' criticism of the idea that the doxastic deliberator possesses an aim is that this aim does not behave like other non-epistemic aims we are familiar with, such as Paul's aim to learn French. Owens expresses this idea by maintaining that the doxastic deliberator's purported aim does not *integrate* with their other aims in the manner we would ordinarily expect of an aim: that is to say, their deliberation

exhibits exclusivity. However, we will now argue that not all ordinary, non-epistemic aims behave in the same way as Paul's aim to learn French. We will explore four different ways in which a person's various non-epistemic aims can interact with one another, and maintain that the fourth of these provides us with an example of a non-epistemic aim that requires a form of deliberation characterised by an analogue of exclusivity.[16]

Let us remind ourselves of Paul's aim to learn French. As we have seen, in so far as Paul is rational, his aim to learn French integrates with his other aims. He both aims to learn French and to hold down his job in England, and thus enrols in French evening classes in England. The fact that Paul does not pursue the *optimal* strategy for realising his aim to learn French – moving to France – does not threaten our attribution to him of the aim to learn French. Rather, we simply recognise that Paul has other aims as well, which are *compatible* with his aim to learn French when all of these aims are pursued *suboptimally*. This, of course, is not to say that Paul could not value learning French so highly that it would be rational for him to move to France. It is simply to say that in so far as he is rational, whatever his aims and the various weights he attaches to them *are*, he will integrate them accordingly.

Now, as Owens says, the guesser's aim integrates with their other aims in this manner. The quiz contestant integrates their truth aim *qua* guesser with their aim to win the prize money in the sense that they will be moved to have a guess at the right answer even without sufficient evidence to guarantee truth in order to be in with a chance of winning the prize. In so doing, they pursue both of their compatible aims in a suboptimal manner.

However, this is not so with all possible aims, as Steglich-Petersen (2009) points out. Consider two aims whose pursuit is *irrelevant* to one another. As Steglich-Petersen says, one might wonder how any two aims can ever *really* be irrelevant to one another as the pursuit of any given aim will make its demand on a person's finite resources, thus impacting upon the pursuit of another. Nonetheless, as he says, we can certainly identify cases in which two aims *seem* completely irrelevant to one another in the context of ordinary deliberation. Consider Paul's aim to attend his French class this evening and his aim to take his children to their swimming lesson

---

[16] Steglich-Petersen (2009) also identifies the first, second, and fourth of these.

during the day two weeks on Saturday. His deliberation regarding the one aim is practically irrelevant to his deliberation concerning the other. Hence, one need not integrate the pursuit of these two aims. However, one might object that in so far as one is to count as pursuing either of them, one need integrate it with other aims one has that *are* relevant to its pursuit. Nonetheless, this does not always mean that one need pursue an aim suboptimally, which brings us to our third category of interaction between aims. These are what we might call *killing two birds with one stone* cases: cases in which two aims are not irrelevant to one another – so the achievement of one does have *some* bearing upon the achievement of the other, but, nonetheless, one need not compromise with respect to the means one takes to achieve either aim. Imagine, for example, that some French children attend the same swimming class as Paul's children and that during the swimming lesson, Paul can make conversation in French with their talkative and grammatically precise father. In such a case, Paul's aim to take his children swimming and his intention to learn French are both optimally realised (or contributed to in the case of the latter aim) by his attendance: fortuitously, no compromise is made regarding either aim.

And, finally, there is a fourth kind of interaction between aims – between aims that are *mutually incompatible*. Imagine Paul's aim to take his children to a swimming lesson between six and seven one evening and his aim to attend his French class during *precisely* that same hour. Paul cannot achieve both of these aims and so one of them will have to be relinquished. Steglich-Petersen points out that this is exactly how the doxastic deliberator's aim behaves with respect to all of their other aims: they have an aim that is mutually incompatible with all of their other aims. And, as he puts it:

"While it is reasonable [for Owens] to claim that it is essential to something being an aim that it can meaningfully be said to interact with and be weighed against other aims, we should not exclude the possibility that sometimes, due to contingent *de facto* incompatibility such weighing is of the kind that results in discarding aims rather than combining them in action. Weighing which of one's aims is more important is a fundamental and important element in practical deliberation." (Steglich-Petersen, 2009, p. 401)

He compares the doxastic deliberator's aim to the non-epistemic aim:

"…of trying out a new restaurant if and only if it receives good reviews. When subsequently deciding whether to try the restaurant or not, only one consideration will be relevant as far as achieving the aim goes, namely whether or not it received good reviews." (Steglich-Petersen, 2009, p. 402)

His idea is that in the context of doxastic deliberation, one deliberates about *whether to believe that p* and in so doing, one excludes aims other than one's doxastic aim – such as whether they desire to believe that *p*, for example – on the basis that they are incompatible with one's truth aim, just as any consideration other than whether the restaurant received good reviews is excluded as incompatible with the aim of trying out the restaurant *if and only if* it receives good reviews.

Now, one might notice that we have exchanged Owens' talk of the *integration* of aims with talk of aims *interacting*, and of their being *weighed* against one another. One might wonder whether this is illegitimate – perhaps aims that are mutually incompatible cannot be thought of as *integrating* with one another for this very reason. Indeed, the word *integration* does have connotations of intermingling that the word *interaction* does not. However, of course, whatever terminology we use, what matters is whether or not cases of aims that are mutually incompatible with all of a person's other aims do occur in ordinary, non-epistemic contexts. If they do, then it seems that there is no difficulty in understanding the doxastic deliberator's aim as a genuine aim. It is not mysteriously insulated from all of their other considerations in a manner in which no other aim is.

So, the potential example we have of a non-epistemic aim which behaves just as the doxastic deliberator's aim does is:

(T): To try the new restaurant if and only if it receives good reviews.

It seems as though both the restaurant deliberator and the doxastic deliberator are unable to realise their aims in a manner that involves explicitly taking anything other than one factor into consideration: each has an aim that *cannot* be realised in a suboptimal manner by integrating it with their other aims.

However, McHugh (2011a) objects that things are not quite so simple. He asks us to imagine a case in which there is some *uncertainty* as to whether or not the restaurant received good reviews. Suppose that the restaurant deliberator has not read the

reviews, but has seen the facial expressions of people reading them and this evidence suggests that the reviews are good, but there is of course still some doubt. McHugh maintains that in such a scenario, the aim of trying the new restaurant if and only if it gets good reviews would be rationally integrated with one's other aims in the way in which Paul integrates his aim to learn French with his aim to hold down his job in England: the aim will be pursued suboptimally in combination with the deliberator's other aims. For example, the restaurant deliberator's aim could be integrated with their additional aims to avoid pointless expenditure and travel as follows: if the restaurant is cheap and close, he will try it; if it is expensive and far away, he will not. In integrating their aim to try the restaurant if and only if it gets good reviews with their other aims, the restaurant deliberator does not *give up* their aim of trying the restaurant if and only if it gets good reviews; rather, they pursue this aim, whilst running the risk of failing to achieve it.

Now the point here is of course not limited to a particular example. Rather, it is quite a general one about cases in which one aims to ∪ *iff p*. Indeed, it seems that aims of this form will always permit of the kind of integration of mutually compatible aims that it appears that the doxastic deliberator's aim does not. And, as McHugh points out, this is because there is always room to *doubt* whether *p*. However, if we return to (T), we will see a way out of this difficulty. It seems that what Steglich-Petersen was *really* envisaging when providing this example of an ordinary non-epistemic aim, was a case in which one resolves not to go to a new restaurant unless *one reads* good reviews of it beforehand. That is, one aims to go to the new restaurant if and only if one is *satisfied* that it has had good reviews. So, so far, we have been considering an aim of the following form:

1. ∪ *iff p*

Instead, let us now consider this alternative kind of aim:

2. ∪ *iff* one is satisfied that *p*

which in the restaurant deliberator's case would be the following aim:

(T*): To try the new restaurant if and only if one is satisfied that it has received good reviews.

Now, aims of type 2., such as (T*), do not seem to allow for the requisite uncertainty to enable the kind of integration McHugh describes as possible in the case of aims of type 1., such as (T). It is at least very difficult to imagine a case in which one is uncertain as to whether or not one is satisfied that *p*. Although there are of course limits on our self-knowledge, it seems difficult to doubt the idea that we have privileged access to our conscious experience of being satisfied that *p*.[17] And, given that (T*) does not allow for the same kind of uncertainty as (T) it seems as though it does in fact behave in a manner very similar to the doxastic deliberator's aim. That is, neither (T*) nor the doxastic deliberator's purported aim integrate with their other aims. Now, at this point, one might worry that the aim we are now considering:

(T*): To try the new restaurant if and only if one is satisfied that it has received good reviews

is – unlike Steglich-Petersen's original suggestion – no longer a non-epistemic aim. Isn't one's satisfaction here *epistemic*? In response, we should note that the aim is to *try* the restaurant if and only if a certain condition holds, and trying out a new restaurant is certainly not an epistemic activity. Thus, it seems that we have met Owens' challenge by providing an example of deliberation involving a non-epistemic aim that exhibits an analogue of the exclusivity that characterises doxastic deliberation.

3.6 A Dilemma: Circularity or Failure of Exclusivity

So, the suggestion we have arrived at is that rather than:

(D) To believe that *p iff p*;

the doxastic deliberator aims:

(D*): To believe that *p iff* they are satisfied that *p*.

But this aim strikes us as *circular*.[18] That is, being consciously satisfied that *p* seems to be identical to consciously believing that *p*. Perhaps, in order to create some space

---

[17] Although it may be possible to later doubt that one had been genuinely satisfied that *p*, it does not follow from this that at the time at which one considers oneself satisfied that *p* there is room to doubt that this is the case: in fact, the phenomenology of satisfaction seems to crowd out any possibility of this sort of critical distance.

[18] Thanks to Lucy O'Brien for suggesting this line of objection.

between the left and right-hand components of (D), we ought to reformulate the doxastic deliberator's aim as follows:

(D**): To believe that *p iff* they are satisfied that they have sufficient epistemic reason in favour of *p*.

If we combine this new aim with an affirmation of discretion, then we can forge a gap between being satisfied that one has sufficient epistemic reason in favour of *p* and being satisfied that *p*, and hence between the former and believing that *p*. It looks as though we would then have a non-circular aim to attribute to the doxastic deliberator that is also subject to exclusivity – just what we were looking for.

However, as we saw in the previous section, we have not been given reason to affirm discretion – we have not been given reason to think as Frankish (2007, p. 541) puts it that:

"…there is a distinction between reasons that are sufficient to permit belief and reasons that are sufficient to compel it."

Previously, we argued that the intuitive appeal of the kinds of examples provided in support of discretion can be at least equally well explained by the truth of contextualism about judging that one has sufficient epistemic reason in favour of *p*. We will now consider a positive reason to deny that discretion holds, which could not fully be explored within the bounds of the present work, but must nonetheless be mentioned. When we introduced Shah's (2003) use of the term *transparency* to describe an aspect of the phenomenology of doxastic deliberation, we noted that this use is not to be confused with the use of the term made in the self-knowledge literature. When discussing self-knowledge, the term is employed to describe the manner in which one comes to know about one's mental life. Now, it seems intuitive that such transparency holds of the process by which one comes to know that one is satisfied that one has sufficient epistemic reason in favour of *p*. One comes to know that one is satisfied regarding one's reasons not by looking into oneself *via* some kind of mechanism of *inner sense*,[19] but rather by looking out into the world and considering one's reasons. As Richard Moran (2001a) puts it, we come to know such

---

[19] David Armstrong (1968) is a key contemporary proponent of such an inner sense theory of self-knowledge.

facts about ourselves by treating a *theoretical question* about whether we are in a certain condition as amounting, from our perspective, to a *deliberative question* – a question the answer to which one *decides* rather than *discovers*. Now, if this is right, then it seems that when one asks oneself whether one is satisfied that one has sufficient epistemic reason in favour of *p*, one goes about answering this question in the same way that one goes about answering the question *whether to believe that p*. That is, on both occasions, all one can do to answer the question is to consider one's reasons. If this is the case, then there would be no room for discretion, as in coming to know that one has sufficient epistemic reason in favour of *p*, one would have already done all one could to come to believe that *p*.

Given this, it seems that our attempt to conceive of the doxastic deliberator themselves as possessing an aim is in trouble. We have tried to model the doxastic deliberator's aim as *per* (D*) on the restaurant deliberator's aim:

(T*): To try the new restaurant if and only if one is satisfied that it has received good reviews.

However, there is a crucial difference between (T*) and (D*). Trying a new restaurant is not identical to being satisfied that it has received good reviews, whereas believing that *p is* identical to being satisfied that *p*. It seems then that we are faced with a dilemma when it comes to characterising the doxastic deliberator's aim. On the first horn of the dilemma, we formulate the doxastic deliberator's aim as follows:

1. ∪ *iff p,*

in which case there is a gap between the left and right-hand side of the formulation, which means that deliberation concerning the aim will fail to exhibit exclusivity. Alternatively, on the second horn of the dilemma, we formulate the doxastic deliberator's aim as follows:

2. ∪ *iff* one is satisfied that *p,*

in which case there is no gap to allow for failure of exclusivity, but the aim is thereby circular. As a result, it seems that we are forced to agree with Owens that it is illegitimate to think of the doxastic deliberator *themselves* as aiming at anything.

3.7 Demandingness and Aiming at Knowledge

However, before renouncing the idea that the doxastic deliberator has any kind of aim, Owens (2000) claimed this aim to be *knowledge*, as opposed to *truth*, and McHugh (2011c) continues to maintain that this is the case. Is there any hope for this option?

Owens (2000) and McHugh (2011c) claim that the best way of explaining demandingness is to maintain that the doxastic deliberator aims at *knowledge*. Their claim is not that the doxastic deliberator does not aim at truth – knowing that *p*, of course, entails that *p* is true – but rather that their most *basic* aim is knowledge. Their idea is that the doxastic deliberator cannot believe that *p* if they take themselves to have insufficient evidence that *p* as to take oneself to have insufficient evidence that *p* is to take oneself to fail to know that *p*. And, one cannot be counted as aiming at knowledge in believing that *p* if one understands oneself as failing to know that *p*. So, the thought is that what it means to understand oneself as having *sufficient evidence* that *p* is equivalent to understanding oneself as having evidence that is good enough to *know* that *p*. Now, of course, this is not to say that one must explicitly characterise one's evidence under this description. We have maintained throughout that in order to count as a doxastic deliberator, one need not explicitly ask oneself the question *whether to believe that p*. And, similarly, one need not explicitly think: *I have evidence sufficient to know that p*, in order to count as a believer. Rather, this can remain implicit: in order to count as a doxastic deliberator, one need simply find oneself unable to believe that *p* without a certain amount of evidence.

However, attempting to explain demandingness in this manner will not work as it falls foul of the second horn of the dilemma we have outlined. It involves a characterisation of the doxastic deliberator's aim in the form of 2. ∪ *iff* one is satisfied that *p*. And, as we have seen – given that discretion is false – we cannot understand this aim as anything but circular: being consciously satisfied that *p* *is* consciously believing that *p*.

So, it seems that we must conclude, as Owens (2003) later does, that the doxastic deliberator themselves does not possess a doxastic aim. Thus, we need an alternative conception of our regulation claim (R), i.e., the claim that if the doxastic deliberator consciously regulates their belief, they do so solely for truth. Instead of conceiving of

the doxastic deliberator as aiming to believe that *p iff* they are satisfied that *p*, it seems that we should conceive of (R) as follows. The belief that the doxastic deliberator forms as the culmination of doxastic deliberation *itself* aims at truth in the sense that coming to hold it for the doxastic deliberator simply *is* coming to be satisfied that it is true. So, when considering *whether to believe that p*, the doxastic deliberator experiences this question as transparent to the question *whether p* because, in so far as they understand the concept of belief, they understand that forming the belief that *p is* reaching a state of satisfaction that *p*.

Now, at this point, one might wonder whether it is legitimate to think of this kind of conscious regulation for truth as *belief's aiming at truth*. Having concluded that the doxastic deliberator themselves does not possess an aim, perhaps we ought to maintain with Boyle (forthcoming a, p. 25) that:

"…believing is neither an aim we pursue nor an activity structured toward an aim."

It seems that the right thing to say here is that beliefs themselves do not literally aim at truth. As Wedgwood (2002, p. 267) puts it:

"Beliefs are not little archers armed with little bows and arrows: they do not literally "aim" at anything. The claim must be interpreted as a metaphor."

Indeed, we have known all along that Williams' (1973) claim was only ever intended as an evocative metaphor. The question we have been wrestling with is how best to spell out this metaphor. And, rather than opting to understand this claim in terms of the believer's conscious aim, we have opted to explain it in terms of a construal of their coming to believe that *p* as their coming to be satisfied that *p* is true. Of course, we have also sought to explain the metaphor in terms of belief's constitutive correctness condition that we will explore further in the next chapter. Before we do so, however, we will now turn to consider the relationship between belief and knowledge a little further.

Although the doxastic deliberator does not aim at knowledge, perhaps demandingness ought to convince us that belief itself aims at knowledge, rather than truth. Indeed, it seems that being satisfied that *p* is taking oneself to know that *p*. However, Williams (1978, pp. 23-31) maintains that there is no distinction between conceiving of belief as aiming at truth and as aiming at knowledge and we are now in

a position to see why this is so. As Williams says, to become consciously satisfied that *p* is *both* to become consciously satisfied that *p is true*, and in so doing, to become satisfied that one *knows that p*. Thus, to say that beliefs aim at truth in this sense is to say that they aim at knowledge.

Let us now briefly explore the nature of this relationship between belief and knowledge. In support of the idea that to say that beliefs aim at truth is to say that they aim at knowledge, let us begin by observing that it would be strange to understand oneself as believing that *p*, and yet to consider from the first-person perspective whether one knows that *p* and answer in the negative. As McHugh (2011c, p. 385) puts it:

 "This conclusion seems already to amount to a withdrawal of belief that *p*, in so far as your mental states are coherent."

But why was there a need to stipulate that one consider the matter *from the first-person perspective*? Well, it appears that, although demandingness is true, there is no bar to one's taking a third-person perspective on one's situation and realising that one may be wrong to believe that *p*. From such a perspective, one can realise that *for all the evidence one has*, *p* might be false. This is the only sense in which Harold Pritchard (1967, p. 61) is correct when he claims that:

"We must recognize that whenever we know something we either do, or at least can, by reflecting, directly know that we are knowing it, and that whenever we believe something, we similarly either do or can directly know that we are believing it and not knowing it."

Indeed, it seems that this fact about this kind of third-person reflection upon our own beliefs explains one ordinary use of the word *believe*, where it signifies a lack of full assurance. Imagine a case in which one says *I believe it is on the kitchen table* as opposed to simply *It is on the kitchen table*. One might think that such a case constitutes evidence *against* the claim that beliefs aim at knowledge. After all, the speaker is expressing doubt by their very use of the word *believe*. However, in so far as this represents a genuine claim to believe that *it is on the kitchen table* as opposed to something like *I seem to remember it's being on the kitchen table*, or *there is a chance that it is on the kitchen table*, the person speaking is simply expressing their

awareness that although they are satisfied (and hence take themselves to know) that it is on the kitchen table, they are aware that they are not infallible – it nonetheless might not be on the table. It is the fact that *believe* is often used to express fallibility in this manner that, it seems, explains why its use can also constitute a form of politeness when used sincerely or irony when used insincerely as in the locution *I believe you'll find….*

It is this ability to switch between first and third-person perspectives on one's beliefs that is played upon to generate the famous Paradox of the Preface. The Paradox of the Preface arises when an author apologises in the preface for the inevitable errors contained within the book. Given that the author is to be understood as asserting both that the book contains errors, as well as every claim that is made within the main body of the work, a puzzle is generated. There is much that could be said about this puzzle, but here we shall simply observe that it is a switch between first and third-person perspectives on the part of the author that enables them to make both their claim in the preface and each of the individual claims in the book coherently. From within a first-person perspective, for each claim in the book p, they are satisfied that p (and hence take themselves to know that p).[20] However, when adopting a third-person perspective upon their endeavours in researching the book, the author realises that for any one claim (or indeed all of the claims – though they might think this unlikely), they may be wrong: the claim may be false.

But doesn't this ability to switch between first and third-person perspectives leave the believer in a somewhat *unstable* position? That is, once one adopts a third-person perspective and realises that despite one's evidence it may be the case that $\sim p$, wouldn't one bring this realisation back with one into the first-person perspective and hence undermine one's belief that p? This is in fact how sceptical worries set in, when they do. The realisation that for all one knows one might be dreaming, if held on to, does undermine all of one's beliefs. Nonetheless, the fact that it is very difficult (and certainly pathological) to hold on to the idea that such a hypothesis may turn out to be true seems to evidence our marvellous ability to insulate thoughts from a first-person perspective about whether to believe individual propositions from

---

[20] They may of course opt to write things in their book that they are not totally certain they believe, but if this is the case, their apology in the preface loses its puzzling nature.

such sceptical hypotheses, which can only really be entertained from a third-person perspective.

This brings us to the question as to whether there is anything more that we can say about being satisfied that p than that it involves taking oneself to know that p. Don't we owe an explanation as to how *much* epistemic reason one needs to reach such a state of satisfaction in a given situation? We will return to the issue as to how much evidence is sufficient evidence in the next chapter, but let us simply note for now that to provide a fully informative and perhaps reductive account of belief, an answer to this question would need to be forthcoming. But, an answer to this question would constitute an account of what knowledge is, which, thankfully, is beyond the scope of this thesis. For now, let us leave the terms *satisfied* and *knowledge* as unanalysed place-holders in a full account of the nature of the relationship between belief and knowledge.

Conclusion

We had arrived at the end of the last chapter at the idea that beliefs aim at truth in the sense that if they are consciously regulated, this regulation is solely for truth. In this chapter, we have explored how we should understand this claim. In particular, we have considered whether we should conceive of it as involving the attribution of an aim to the doxastic deliberator themselves. We began by considering Owens' (2003) argument that the doxastic deliberator themselves does not aim at anything, because their purported *aim* does not *integrate* with their other aims in the manner we would expect of an ordinary aim: that is, doxastic deliberation exhibits exclusivity. We explored the option of responding by denying that this is so. However, we saw that the cases put forward in an attempt to establish this conclusion were not convincing as our response to them could equally well be explained by our sense that whether one is satisfied that p will be determined by the context. We then attempted to construct an example of a non-epistemic aim that carries with it a form of deliberation subject to a non-epistemic analogue of exclusivity. We saw that we would be unable to do so on the model of 1. U *iff p* because, as long as one can be uncertain that p, exclusivity fails. We then turned to the model of 2.U *iff* one is satisfied that p. It seems that non-epistemic aims of this kind do carry with them a non-epistemic form of exclusivity. However, we observed that since discretion is

implausible, this aim would be circular in the case of the believer: being consciously satisfied that *p is* consciously believing that *p*. We were thus faced with a dilemma. On the first horn we conceive of the doxastic deliberator as aiming to believe that *p iff p*. Deliberation concerning any such aim will fail to exhibit exclusivity in so far as there is room for uncertainty as to whether *p* obtains. However, on the second horn of the dilemma, when we attempt to secure exclusivity by thinking of the doxastic deliberator as aiming to believe that *p iff* they are satisfied that *p*, our aim is a circular one because being consciously satisfied that *p* is consciously believing that *p*. We argued that conceiving of the doxastic deliberator as aiming at knowledge falls fouls of the second horn of the dilemma as taking oneself to know that *p* is being satisfied that *p*. Thus, we concluded that we ought not to understand the doxastic deliberator themselves as aiming. Rather, we argued, beliefs aim at the truth in the sense that the believer is disposed to consciously regulate their belief for truth such that they will not be able to form a belief that *p* unless they are satisfied that *p* – because consciously believing that *p* is being consciously satisfied that *p*. This understanding of belief captures both the exclusivity and the demandingness of doxastic deliberation. We defended the legitimacy of conceiving of this as belief itself as aiming and we then saw that understanding belief itself to aim at truth in this way is equivalent to understanding it as aiming at knowledge, because being satisfied that *p* is also understanding oneself to know that *p*. We closed by briefly exploring this connection between belief and knowledge.

4. The Correctness Condition

Having come to a better understanding of the regulation condition, it is now time to turn to the correctness condition, which we have argued underpins the regulation condition. In chapter two, we saw that the transparency of doxastic deliberation is ultimately to be explained by the fact that the doxastic deliberator understands that their belief that *p* is only answerable to epistemic norms, and that this means that their belief that *p* will be *correct* if and only if *p*. However, the doxastic deliberator does not have direct access to whether or not *p* is true, as we emphasised in the previous chapter. Thus, their deliberation as to *whether to believe that p* is governed only *indirectly* by the correctness condition, *via* derivative epistemic norms that provide them with reliable means of satisfying the correctness condition. The correctness condition tells us when a belief is correct – namely when it is true – and these subjective epistemic norms tell us when a belief is rational. In this chapter, we will argue that the subjective norms that govern doxastic deliberation are what we will call *normative exclusivity*, and *normative demandingness*. These are the normative analogues of descriptive exclusivity and demandingness, as introduced in the previous chapter. According to normative exclusivity, one *ought* to believe that *p* only if one does so solely on the basis of objective epistemic reason in favour of *p*. In addition, according to normative demandingness, one *ought* to believe that *p* if and only if one does so on the basis of *sufficient* objective epistemic reason in favour of *p*.[21] These prescriptions combine to yield:

(Ra): One should believe that *p* if and only if one does so solely on the basis of sufficient objective epistemic reason in favour of *p*.

We will begin by briefly considering what it *means* to call a true belief a *correct* belief. We will then outline both normative exclusivity and demandingness in greater detail and argue that they can be derived from the correctness condition. Having fully articulated our understanding of these epistemic norms in this way, we will consider two potential lines of objection to the idea that they are constitutive of belief. First, we will consider *pragmatism*, which we will understand, following Owens (2000), to

---

[21] As with descriptive exclusivity and demandingness, normative exclusivity and demandingness pertain to all forms of conscious belief regulation – not merely formation. Unlike the descriptive forms of these claims, however, the normative versions pertain to all belief regulation – be it conscious or unconscious. As in the previous, we will restrict our focus, for the sake of simplicity, to belief-formation in the context of doxastic deliberation.

be the claim that belief is answerable to practical norms, the most fundamental of which has it that an action is (practically) rational just if the occurrence of that action (along with its consequences) would be desirable. We will argue that we can explain the intuitions the pragmatist invokes by acknowledging that, although beliefs themselves are never answerable to anything other than epistemic norms, it may nonetheless be practically rational to attempt to foster certain beliefs in oneself by putting oneself in certain situations, for example. This rationality pertains to the activity of fostering rather than the resultant belief, which will always be irrational if its formation contravenes normative exclusivity and demandingness. The second potential objection to the idea that belief is answerable to the correctness condition – and hence normative exclusivity and demandingness – we will consider is the notion that we do not possess the requisite kind of *control* over our beliefs in order to be held responsible for them in accordance with such norms. However, we will argue that we have what we will call *reflective control* over our beliefs – roughly, we can control our beliefs insofar as we can reflect upon our reasons for holding them – which renders the manner in which we hold one another (and ourselves) responsible for our beliefs coherent. In order to establish this claim, we will respond to Owens' (2000) critique of the reflective control model, according to which pragmatic considerations play a role in rationalising our beliefs that precludes reflective control. We will argue that his account does not rule out a credible alternative account of justification which extends no role to pragmatic considerations in the rationalisation of belief, at least in paradigmatic cases. We will close by articulating in further detail the manner in which reflective control explains our practices of holding one another responsible and the precise way in which the role of judgement in reflective control is to be understood.

In this chapter, it is particularly important to bear in mind that what we are doing is systematising our folk concept of belief. For example, we will argue that the correctness condition is constitutive of our concept of belief, but in so doing, we do not commit ourselves to the existence of normative facts. We are merely claiming that this is what our concept of belief involves. Whether or not this concept will turn out to refer in its entirety is then a further question.

4.1 *Correctness* as a Term of Art

It is important to emphasise that referring to a belief as *correct* is to employ a term of art. In ordinary parlance, if one were to call a particular belief *correct*, it would be unclear what one meant. In particular, one might wonder whether the belief was being asserted to be *true*, or *rational*. We have been using the term *correct* to indicate a true belief but, of course, in claiming truth to be belief's standard of correctness, it need not be denied that beliefs can be evaluated along additional dimensions. Indeed, as we will presently argue, we can derive subjective epistemic norms, which tell us when a particular belief is rational, from the objective correctness condition. Before we do so, however, let us have a slightly closer look at what it means to call a true belief correct. This claim can be interpreted variously,[22] but we will understand it in the manner Wedgwood (2002, p. 267) outlines here:

"'Correct' is not just a synonym for 'true'. To say that a mental state is "correct" is to say that in having that mental state, one has got things "right"; one's mental state is "appropriate". To say that a mental state is "incorrect" is to say that in having that mental state, one has got things "wrong" or made a "mistake"; one's mental state is in a sense "defective". Clearly, there is nothing wrong or defective about false propositions as such; what is defective is *believing* such false propositions. Moreover, other mental states besides beliefs, such as choices or decisions, can also be wrong or mistaken or incorrect. So 'is correct' also does not just mean "is a belief in a true proposition"."

We will also refer to the correctness condition so understood as an *epistemic norm*. Furthermore, we will now argue that it is the fundamental epistemic norm, from which we can derive further epistemic norms that tell us when a belief is rational.

4.2 Normative Exclusivity and Normative Demandingness

In the previous chapter, we argued that the ultimate explanation of transparency lies in the fact that the doxastic deliberator recognises that the correctness condition is constitutive of belief. Now, so far, the only epistemic norm we have considered is the correctness condition itself. However, it seems that we cannot give a complete explanation of transparency in terms of the correctness condition alone, because the

---

[22] See McHugh (2012a).

correctness condition informs us that a belief that *p* is correct if and only if *p* is true, but the doxastic deliberator may not know whether *p* is true. Thus, it seems that their deliberation must be governed by additional epistemic norms that instruct them how to get to the truth.

Now, in the previous chapter, we introduced the phenomena of exclusivity and demandingness. We understood exclusivity to be the claim that the doxastic deliberator *can* only form the belief that *p* if they think of themselves as doing so on the basis of epistemic reasons, and we took demandingness to be the claim that the doxastic deliberator *can* only form the belief that *p* if they think they are doing so on the basis of sufficient epistemic reason in favour of *p*. When we introduced this terminology of exclusivity and demandingness, we noted that both can be given a normative construal in addition to their descriptive one. It is this normative construal of both exclusivity and demandingness that serves as the subjective translation of the correctness condition and informs the doxastic deliberator about how best to go about believing the truth. What we will call *normative exclusivity* is the claim that one *ought* to form one's beliefs solely on the basis of objective epistemic reasons. *Normative demandingness*, on the other hand, is the claim that one *ought* to form one's beliefs on the basis of sufficient objective epistemic reason.

Now, it is important to note that these purported norms are to be understood as regulating the process of belief formation on the basis of the subjective epistemic reasons one already possesses, rather than the process of inquiry, or the gathering of such reasons. Normative demandingness may have implications for the process of inquiry, though. That is, according to normative demandingness, if one has insufficient epistemic reason in favour of *p*, one should not form the belief that *p*. Now, if one finds oneself in a situation in which one does not have sufficient epistemic reason regarding a certain proposition, *p*, and one desires to make up one's mind about *p*, this will mean that it will be rational for one to seek further epistemic reasons pertaining to *p*, but the influence of normative demandingness on inquiry can only be indirect in this way.

Now, one might wonder whether it makes sense to think of the doxastic deliberator as subject to normative exclusivity and demandingness, precisely because their deliberation exhibits *descriptive* exclusivity and demandingness. That is, they cannot

think of themselves as doing anything other than forming their belief solely on the basis of sufficient epistemic reason, so what sense is there in informing them that they should form their beliefs in this way? However, even though the doxastic deliberator can only understand themselves to be forming the belief that *p* solely on the basis of sufficient objective epistemic reason in favour of *p*, they can nonetheless fail to do so, as we have seen. The doxastic deliberator forms an *irrational* belief that *p* when they *think* that they do so solely on the basis of sufficient objective epistemic reason in favour of *p* yet this is in fact *not* the case. Sometimes their belief formation is influenced outside of consciousness by what they want to believe, for example (psychologists call this *hot* biasing), and sometimes they are simply inattentive to their reasons (an example of *cold* biasing). Furthermore, normative exclusivity and demandingness pertain to *all* beliefs, not simply those formed by the doxastic deliberator – and beliefs can be formed outside of conscious attention without adherence to normative exclusivity and normative demandingness. Given that, even in the context of doxastic deliberation, the influence of non-epistemic factors always threatens, the prescription to guard against these factors, embodied in normative exclusivity and demandingness, is *prima facie* coherent. Even if one cannot but think of oneself as having formed a belief solely on the basis of sufficient epistemic reason, normative exclusivity and demandingness counsel one to be attentive to one's epistemic objective reasons because, whatever one thinks one has done, one will have exhibited a failing if one's belief is not formed solely on the basis of such reasons. In the final sections of this chapter, we will consider in some detail the kind of control one must have over one's beliefs in order to be considered responsible for them in the manner that would be required for normative exclusivity and demandingness to hold.

A second potential objection pertains to the relationship between normative exclusivity and demandingness and the correctness condition. Above we have claimed that the former can be *derived* from the latter, but one might object to this claim for the following reason. The correctness condition holds that a belief is correct if and only if it is true, but it says nothing about how one must arrive at that belief. So, even if a belief is grossly irrational, if it is true, it is correct, according to the correctness condition. However, even though irrational beliefs can be correct, the only way in which it makes sense for the doxastic deliberator to consider *whether to*

believe that p (understanding the correctness condition to be constitutive of belief) is to consider what it would be *rational* to believe. Rational belief is the best *means* they have to get to the truth. In this connection, Wedgwood (2002, p. 282) claims that:

"The following principle seems a plausible claim about norms in general (not just epistemic norms). If there is a fundamental norm that directs one to achieve a certain outcome and that outcome is an end that one can achieve only by using means to that end, then there is also a secondary norm that directs one to use means that it is rational for one to believe to be sufficiently reliable means to that end."

And, this seems to be exactly how it is with normative exclusivity and normative demandingness. Combining these two epistemic norms we get the following rational constraint upon a believer:

(Ra): One should believe that *p* if and only if one does so solely on the basis of sufficient objective epistemic reason in favour of *p*.

The explanation as to why this prescription holds does seem to be that it is rational for one to believe that following the prescription will be a sufficiently reliable means to the end of having a true (and hence correct) belief.[23]

4.3 Evidentialism *vs*. Pragmatism

However, at this point, one might object that it is not so obvious that belief *is* subject solely to such epistemic norms. Indeed, there is a long-standing debate between so-called *evidentialists* and *pragmatists* that is relevant here, but this debate has not always been sensitive to the distinction we have drawn between normative exclusivity and normative demandingness. Thus, in order to engage with this debate through the lens of the distinctions we have developed, we should first distinguish between three kinds of evidentialism – evidentialism that affirms normative exclusivity, evidentialism that affirms normative demandingness, and evidentialism

---

[23] One might wonder whether in the previous chapter we ought to have considered the idea that the doxastic deliberator aims to form a rational belief. However, such an aim would fall foul of the second horn of our dilemma: the doxastic deliberator would be aiming to believe that *p iff* they were satisfied that they were doing so in accordance with (Ra), which amounts to aiming to believe that *p iff* they were satisfied that *p*. Such an aim would be circular, as we have argued.

that affirms both. Evidentialism is often classified as essentially a thesis about what *kind* of reasons ought to influence belief as *per* normative exclusivity:

"Evidentialism is the thesis that what justifies belief in *p* is just evidence in *p*'s favour." (Owens, 2000, p. 23)

But it is also often thought of as an affirmation of normative demandingness as well. In the words of perhaps *the* most famous evidentialist, William Kingdon Clifford (1999, p. 77):

"It is wrong always, everywhere, and for anyone to believe anything on *insufficient* evidence." (italics added)

And, of course, an evidentialist could affirm both of these theses, maintaining what we have called (Ra). Pragmatism is essentially the claim that the norms that govern action – what we will call *practical norms* – also govern belief regulation. Given this, it is characteristic of the pragmatist to deny normative exclusivity, maintaining instead that it is often the case that what one ought to believe is in part determined by non-epistemic reasons. Furthermore, some pragmatists will hold that sometimes one ought to believe that *p*, even if one has insufficient epistemic reason to believe that *p* (*contra* normative demandingness). An arch-pragmatist, James (1931) denies both normative exclusivity and normative demandingness when he claims that:

"Our passional nature not only lawfully may, but must, decide an option between propositions, whenever it is a genuine option that cannot by its nature be decided on intellectual grounds."[24] (James, 1931, p. 11)

James (1931) invokes the famous example of Pascal's Wager in support of his pragmatism. Pascal (2006) notoriously maintains that if we perform a cost-benefit analysis, it becomes clear that one *ought* to believe that God exists. We can reconstruct Pascal's thinking as follows. He observes that there are four main situations one might find oneself in with respect to belief in God:

1.   Believe in God and God exists;

---

[24] Both Clifford and James slip between what appear to be epistemic, prudential, and even moral uses of the term *ought* when discussing what we ought to believe. We shall set the question as to the morality of belief regulation aside.

2.   Believe in God and God does not exist;

3.   Do not believe in God and God exists;

4.   Do not believe in God and God does not exist.

Now, presuming that belief in God is the sole determinant as to whether one is sent to heaven or hell and that there are no heaven and hell if God does not exist, 1. is the best situation in which to find oneself: one is rewarded for one's belief by an eternity of bliss in heaven. 4. is the next best situation in which to find oneself: one does not have to bother with the sacrifices of belief in God during this lifetime, and will face no punishment for this in the hereafter. 2. is the next best situation: one wastes time in this life worshiping a God who turns out not to exist but then suffers no further in the afterlife. However, 3. is an unspeakably bad situation in which to find oneself – far worse than any of the other eventualities – as one is punished for one's lack of belief in this life by an eternity of torment.

Given this line of reasoning, Pascal entreats us to attempt to foster belief in God by throwing ourselves into a religious life. Now, as we saw in the previous chapter, both descriptive exclusivity and descriptive demandingness are true of doxastic deliberation. This is to say, the doxastic deliberator cannot understand themselves to form a belief that p on the basis of any non-epistemic consideration and they cannot form a belief that p unless they are satisfied that they have sufficient epistemic reason in favour of p. Nonetheless, let us grant for the sake of argument that they are able to undertake a programme of self-manipulation – they attend mass every day – such that they come to form the belief that p. Let us imagine that attending mass in this way enables them to overlook the reasons they have against belief in God and that they somehow manage to think of their belief in God as unrelated to their initial non-epistemic reasons even though it continues to be so.[25] The question we are faced with here is whether this belief is *rational*: whether they *should* believe it.

It seems that we can explain the pull of Pascal's thinking by distinguishing between the rationality of an action designed to induce a belief (which *is* governed by practical norms), and the rationality of the belief itself (which is not), as Owens (2000, p. 29) suggests. Although it may be practically rational to attempt to induce a

---

[25] We will return to the topic of self-deception in the next chapter.

belief in God in oneself, whether the belief itself is rational depends solely upon whether its regulation adheres to epistemic norms. We can produce no particular argument for this claim other than by offering it as an explanation of the apparent pull of cases like Pascal's, such as it is, and relying upon the robust intuition that the rationality of belief itself cannot be subject to practical norms.

Now, the pragmatist may acknowledge that there is a distinction to be drawn between whether it is practically rational to attempt to get oneself to form a certain belief and whether the belief itself is (epistemically) rational. Nonetheless, they may charge that this distinction does nothing to help us from a deliberative point of view. The question *should I believe that p* can be given only one all-things-considered answer. Being told that there is a distinction between the status of a belief as epistemically rational, and the status of a belief as the outcome of a practically rational course of action, only puts off the question as to *whether to believe that p*. We need to know whether our answer to this question should advert to considerations pertaining to epistemic or practical rationality, and the pragmatist charges that there are many cases in which it is to practical, and not epistemic rationality that we should turn. However, we can respond to this by maintaining that considerations of practical and epistemic rationality are *not* pertinent to the *same* question, after all. The question to which practical considerations are relevant, and which the pragmatist erroneously takes the doxastic deliberator to be asking, is *should I try to get myself to belief that p*. This is a different question to the question *should I believe that p*, and this latter question is answerable solely to epistemic norms. Thus, we do have the means to settle on an all-things-considered answer to the question *should I believe that p*. This question is only answerable to epistemic norms, even if the all-things-considered answer to the question *should I get myself to belief that p* is answerable to practical norms.

David Papineau (forthcoming, p. 4) puts forward a different kind of example designed to provoke pragmatist intuitions:

"Suppose some quite untreatable form of cancer is common in John's family and that he indeed has it. There is a simple enough test, but John doesn't take it, because he is confident that he doesn't have the cancer. (He feels great!) As a result, he avoids the distress and unhappiness that would be occasioned by his learning the truth."

We might call such cases *Ignorance is Bliss* cases. Given that the cancer is untreatable, Papineau (forthcoming, p. 4) maintains that John "has done nothing wrong at all"; in fact, he cites the case as one in which there is "a positive value" in John's having a false belief. Does this kind of case conflict with our claim that belief is answerable to the correctness condition, and, derivatively, normative exclusivity and demandingness? No: again, it is perfectly consistent with our position that it may be practically rational for John to avoid believing that he has cancer. Nonetheless, in so far as he forms the belief that it is not the case that he has cancer on the basis of insufficient epistemic reason, his belief itself is rationally criticisable. This allows us to accommodate our unease at saying, as Papineau does, that John "has done nothing wrong *at all*" (italics added) in believing a false proposition, but also to acknowledge that John may well be better off under such circumstances.

Finally, there are what we might call *Uninteresting Truths* cases, in which pressure is applied directly to the correctness condition itself. These are cases in which the true belief in question is of utterly no interest. Papineau cites the example of his having a belief about the number of blades of grass on his lawn. Whiting (2012) puts forward the example of having a belief about the length and colour of each hair on David Cameron's left arm. Such true beliefs, Papineau claims, are of no value to anyone.

As for *Uninteresting Truths* cases, it is consistent with the correctness condition and its derivative norms that there are (perhaps always) many more worthwhile things to do than form a belief regarding the number of blades of grass on Papineau's lawn. Nonetheless, the correctness condition should be understood as claiming that *if* one undertakes to form a belief about such a matter, in believing falsely that there are x blades of grass, one does something *worse* than one would do in believing truly that there are y.

4.4 Deciding to Believe

So far, then, we have defended the idea that we can derive both normative exclusivity and demandingness from the correctness condition, and that we can explain away any intuitions that might appear to support pragmatism by distinguishing between whether it is rational to attempt to acquire or avoid certain beliefs, and whether, after one has considered a proposition, the belief in this proposition is itself rational. We will now consider a second line of objection to the

idea that belief is subject to the epistemic norms we have outlined. This is the idea that we cannot be said to have the requisite kind of control over our beliefs in order to be held responsible for them in the manner required by normative exclusivity and demandingness. This objection will occupy our attention for the remainder of this chapter. In the present section, we will consider one brand of response to the objection, along the lines of the doctrine known as *voluntarism*. We will distinguish two versions of voluntarism, and argue that neither position is able to account for the sort of control that would be required for epistemic norms to be regarded as applying to doxastic deliberation.

A natural thought is that in order for a norm to apply to someone, they need to be *free* to comply with it. A further natural thought is that what it means to be able to freely comply with a norm is to be able to *decide* to do so. However, Williams (1973) famously claims that one cannot *decide to believe*. Now, as Owens (2000, p. 78) points out, the relationship between decision and voluntariness is complex:

"'Decision might refer to a practical judgement – the judgement that I *should* raise my hand. On the other hand, 'decision' could refer to an executive event, the formation of an intention, a state which ensured that I will perform the act decided upon.''

Thus, we may distinguish two versions of the claim that one can decide to belief. The first, which we will call *judgement-based voluntarism*, is the idea that one can make a practical judgement – on the basis of one's practical reasons – that one ought to form a belief, and consequently form a belief on this basis. The second version, which we will call *will-based voluntarism*, is the idea that, by a sheer act of will, one can come to *believe that p*.[26] We will see that Williams' objection is sustained in relation to both of these positions.

Let us begin with judgement-based voluntarism. In the previous chapter we saw that, from the perspective of the doxastic deliberator, their beliefs are constrained by descriptive exclusivity and descriptive demandingness, so they cannot adopt a belief on the basis of a decision to do so, *irrespective of these constraints*. But what if these

---

[26] It should be noted that these two positions are often conflated under the banner of *voluntarism*. In particular, see Frankish (2007). Nonetheless, as we will see, these two positions *should* be held apart, because they involve substantively different conceptions of what it is to make a decision.

constraints are taken into consideration? That is, what if, as Frankish (2007) puts it, the *enabling condition* of descriptive demandingness has been met, and one can decide to believe that *p* under such circumstances without violating descriptive exclusivity? Frankish thinks that this is possible. He claims that once one takes oneself to have sufficient epistemic reason in favour of *p*, there is still room for an act of belief formation, which he calls *endorsement*. Furthermore, he claims that, although:

"…pragmatic considerations cannot determine *which side* of the fence we come down on; that will be determined by the evidence available to us… [nonetheless,] they may determine *that* we come off the fence at all – that we form a belief on some matter rather than withholding judgement." (Frankish, 2007, pp. 541-542)

So, Frankish's idea is that one can make a practical judgement in favour of adopting a particular belief (an endorsement), without violating either descriptive exclusivity or demandingness.

However, as we have seen, the idea that once one understands oneself to have sufficient epistemic reason in favour of *p*, one need not yet believe that *p* is implausible: discretion, and, along with it, judgement-based voluntarism, ought to be denied.

This brings us to will-based voluntarism. Again, descriptive exclusivity and demandingness clearly rule out the possibility of forming the belief that *p* on the basis of a sheer act of will. What about the possibility of meeting these conditions, and *then* deciding to believe through an act of will? This proposal runs into exactly the same difficulties as the proposal considered above. Given the falsity of discretion, the imagined situation, in which one has sufficient epistemic reason to believe that *p*, but needs to propel oneself "off the fence" through an act of will, is not a situation that could obtain.

4.5 Indirect Control

Given that such voluntary control over our beliefs is to be ruled out, it might be suggested that we are subject to epistemic norms such as normative exclusivity and demandingness because we have *indirect* control over our beliefs. Perhaps I honour epistemic norms insofar as *I put myself in a position* to adhere to them. For example,

I seek to learn about resuscitation procedures and hence, in order to adhere to epistemic norms, I enrol myself in a first-aid course. As Galen Strawson (2003, pp. 231-3) describes, perhaps:

"[T]he role of genuine action in thought is at best indirect. It is entirely *prefatory*, it is essentially – merely *catalytic*. For what actually happens, when one wants to think about some issue or work something out? If the issue is a difficult one, then there may well be a distinct, and distinctive, phenomenon of setting one's mind at the problem… No doubt there are other such preparatory, ground-setting, tuning, retuning, shepherding, active moves or intention initiations. The rest is waiting, seeing if anything happens, waiting for content to come to mind… There is I believe no action at all in reasoning and judging considered independently of the preparatory, catalytic phenomena just mentioned, considered in respect of their being a matter of specific content-production or of inferential moves between particular contents."

However, this cannot be considered to be an adequate construal of the kind of responsibility we take ourselves to have for our beliefs. As Boyle (forthcoming a, p. 16) says, it seems that if someone has done all they can to put themselves in a position to conform to these epistemic norms and yet they have failed to do so, we should regard them as "unfortunate", and this is certainly not how we regard someone whose beliefs fail to conform to epistemic norms: we hold them directly responsible for their belief.

4.6 Reflective Control

Given the failure of voluntarism as well as the indirect construal of epistemic control, it has seemed to many that the most promising option open to one who seeks to maintain that we have the appropriate kind of control over our beliefs to be subject to both normative exclusivity and demandingness is what Owens (2000, p. 3) has called *reflective* control. This is the idea that one is free to exercise control over one's beliefs in so far as one has the capacity to *reflect* upon one's epistemic reasons for holding those beliefs and thereby regulate one's beliefs. This idea has found expression in recent authors such as Christine Korsgaard (1996a), and McDowell (2009). We can think of the idea this way. Just as the compatibilist in the free will debate is correct to maintain that we are free to act insofar as we are free to reflect upon our practical decisions and act in accordance with these, so the right thing to

say about doxastic freedom is that we are free to believe as we please, insofar as we are free to reflect upon our epistemic reasons and believe accordingly. To ask for anything more would be to ask for the wrong kind of freedom, in both cases.[27]

However, Owens mounts a powerful attack on the idea that we have reflective control over our beliefs. He affirms both what we have been calling descriptive and normative demandingness. That is, he maintains that the doxastic deliberator *can* only form the belief that *p* if they take themselves to have sufficient epistemic reason in favour of *p* and that they *should* only form the belief that *p* on the basis of sufficient epistemic reason in favour of *p*. However, even though he affirms what we have been calling descriptive exclusivity, he *denies* normative exclusivity. So, although he thinks that we can only consciously understand ourselves as forming a belief solely on the basis of epistemic reasons, he does not think – *contra* normative exclusivity – that a belief is only rational if it is formed solely on the basis of epistemic reasons. Why? Because he thinks that a full description of what it *is* to have sufficient epistemic reason in favour of *p* will necessarily invoke pragmatic considerations such as the importance of the matter, and the amount of time one has etc. Even though the doxastic deliberator can only take themselves to be forming a belief on the basis of epistemic reasons, their:

"…awareness of inconclusive evidence plays an essential non-reflective role: it combines with a sense of the constraints on… [their] cognitive resources to produce the impression of a conclusive reason." (Owens, 2000, p. 35)

Thus, we have all of the ingredients for denying the idea that we have reflective control over our beliefs. According to Owens' picture, although we need to invoke pragmatic considerations to render a belief rational (*pace* normative exclusivity) and indeed to be able to form a belief, we cannot consciously consider ourselves to be forming a belief on the basis of such pragmatic considerations (as *per* descriptive exclusivity). What this means is that we cannot *reflect* upon what are – as a matter of fact – our reasons for belief, and hence exert control over our beliefs. As a matter of

---

[27] There is an important disanalogy with the moral case, however. In the case of practical deliberation, *akrasia* – judging that all things considered one ought to Ս and yet failing to Ս – is always a possibility. However, there is no such thing as *epistemic akrasia* because judging that one ought to believe that *p* simply *is* believing that *p*, as we have argued.

fact, our beliefs are in part based upon pragmatic considerations, but reflection upon these considerations does not enable us to exert control over our beliefs.

One might wonder why, given that Owens emphasises the role of pragmatic considerations in rationalising belief, we did not already consider his account when we considered pragmatism earlier. The reason for this is that, although he denies normative exclusivity, Owens does not conceive of himself as a pragmatist. He distinguishes between what he calls *pragmatic* considerations, such as time constraints, and *practical* considerations, which are concerned with what one desires. Although he thinks that the former rationalise belief, he denies that the latter ever do, and understands this to amount to a denial of pragmatism. Now, we might wonder about this distinction between pragmatic and practical considerations, but for our purposes we can simply grant him the distinction, and with it the claim that his position is distinct from pragmatism.

Owens argues that, as reflective control was the best hope we had for some kind of control over our beliefs, and, as it turns out that we cannot exercise reflective control over our beliefs, we must give up the idea that we can exercise control over our beliefs. However, although he denies normative exclusivity, he does not conclude from what he takes to be the failure of reflective control that we should not be held responsible to normative demandingness, which he affirms. Rather, he maintains that, as he puts it:

"Some norms are not there to guide action, to govern the exercise of control: their function is to assess what we are." (Owens, 2000, p. 126)

Now, we shall not dispute that this is the case. However, it is worth noting that understanding our responsibility to epistemic norms on this model would amount to a revisionist account of such responsibility. As we have previously emphasised, we are held *directly responsible* to epistemic norms. That is to say, when we accuse someone, for example, of having an irrational belief, we do not criticise them in the manner we would if we were merely criticising them on the basis of the kind of person they are. Rather, when we ask them, *why do you believe that*, we implicitly assume that they could do something about their belief. Thus, modelling our epistemic responsibility in the terms Owens proposes would involve viewing our ordinary practices as based on a significant degree of error. To the extent that we are

motivated to vindicate these practices in our account of epistemic responsibility, therefore, there should be a presumption against accepting Owens' account. In what follows, we will be concerned to cast doubt upon Owens' denial of reflective control, thus showing that he has not demonstrated the need to accept his account, and the theoretical costs that come with it.

It will be useful to begin with a fuller statement of the structure of Owens's position. For any given belief, there is an indefinite number of alternative hypotheses that, if true, would show the belief in question to be false. Owens' claim that pragmatic considerations rationalise belief is based on the idea that, for a belief to be rational, all of these alternative hypotheses need to be ruled out. His further thought – surely correct – is that the epistemic reasons reflectively available to a subject cannot be such as to rule out the entire range of these hypotheses. Thus, for example, consulting an encyclopedia of philosophy might rule out the hypothesis that Russell wrote the *Tractatus*, but it will not rule out the hypothesis that *Wittgenstein* was the pseudonym of some devious philosopher, and that this philosopher and not the real Ludwig Wittgenstein in fact penned the *Tractatus*. More radically, it might be thought that *no* reflectively available epistemic reason could rule out the hypothesis that our experience is engineered by a mad scientist. Thus, it would seem that epistemic reasons alone are in principle incapable of rationalising our beliefs. It does not follow from this, however, that all of our beliefs are irrational. To see this, we should distinguish between an epistemic ruling out and a pragmatic ruling out. Let us say that a hypothesis *h* is epistemically ruled out if the truth of a proposition constituting one of one's epistemic reasons would be incompatible with the truth of *h*. A hypothesis *h* is *pragmatically* ruled out, meanwhile, if pragmatic considerations such as the time available make it the case that the subject will have sufficient epistemic reason for some belief without having epistemically ruled out *h*. Owens' idea is that a belief is rational just when the totality of alternative hypotheses have been ruled out. Given that epistemic reasons alone are never capable of ruling out the entirety of such hypotheses, it follows that pragmatic ruling out is a necessary element of rationalisation. Thus, pragmatic considerations play an ineliminable role in making beliefs rational.

Now, a full engagement with this position would extend beyond the scope of the present thesis. Thankfully, however, such an engagement is not necessary in order to

cast doubt upon Owens' position. Given its theoretically unattractive consequence that our ordinary practices of holding one another responsible for the character of our beliefs is in error, there should be a presumption against Owens' account, as argued above. That is to say, the burden of proof is firmly with him, and he will need to provide a knockdown argument for his position in order for us to accept it. Such an argument would need to refute any coherent and *prima facie* credible philosophical alternatives. Consequently, all we need to do to show that Owens has not established his account is to spell out such a position, and show that his account has not ruled it out. Now, we arrive at such a position if we take issue with Owens' starting premise – namely, that a belief is only rational once the totality of competing alternative hypotheses have been ruled out, epistemically or else pragmatically. Take a simple case of a perceptual belief – e.g. *there is a wine glass in front of me* – that is formed on the basis of my visual experience as of a wine glass in front of me. Now, there is a strong intuitive case for claiming that this experience provides me with sufficient epistemic reason for the belief for it to be rational for me to adopt the belief on this basis, even though this reason does not epistemically rule out the possibility, e.g., that I am a brain in a vat. Now, so far, Owens could agree: the *reason* that my visual experience as of a wine glass provides me with sufficient epistemic reason for rational belief in this case, in spite of the fact that the existence of this visual experience is not incompatible with the truth of the brain in a vat hypothesis, is that I pragmatically rule out such a hypothesis. Owens wants to account for my belief regulation by claiming that this regulation is partially responsive to pragmatic considerations such as the time available to me, and the importance of the issue. But we can put pressure on this idea by imagining that I had infinite time available, or that a correct belief concerning the presence of the glass would win me £1 million. Even in such cases, it is hard to imagine what I would do differently in order to settle on a belief as to whether the wine glass is in front of me. Owens would need to say that, in such cases, having suspended all pragmatic considerations, my perceptual experience no longer provides me with sufficient epistemic reason for rationally believing that there is a wine glass in front of me. But this is far from clear. It is true that on this occasion there are a number of alternative hypotheses that are neither epistemically nor pragmatically ruled out; but Owens would have to site some other fact about the case than this in order for the verdict his position demands – that I no longer hold a rational belief – not to beg the question. It seems that all he has left to

appeal to are our intuitions about the present case, but these are far from conclusive. If, however, I *can* be regarded as possessing sufficient epistemic reason to believe that there is a wine glass in front of me in this case, then there is nothing preventing us from holding that in the ordinary case as well, pragmatic considerations play no role in making it the case that my belief that the wine glass is in front of me is based on sufficient epistemic reason, hence rational.

Now, the foregoing was certainly not intended as a conclusive argument in support of the position that we presented. We simply aimed to show that a coherent position that is not entirely without intuitive support would deny Owens' claim that one needs to have ruled out all competing hypotheses – whether epistemically or pragmatically – for one's belief to be rational. Since Owens' account, as it stands, does not amount to a knock-down refutation of such a position, and since we should require him to refute all competing accounts in just this way given the theoretically unattractive consequences of his own position, it follows that we have not yet been given sufficient reason to accept his account.

Now, Owens claims that our beliefs are always in part caused by pragmatic considerations. Thankfully, such considerations, as well as causing beliefs, partly rationalise them. In the above, we have put pressure on the idea that such considerations do in fact rationalise beliefs, but this does not of itself show that they do not cause our beliefs. If this were the case, then, not only would we not possess reflective control, but our beliefs, being in part caused by considerations that do not rationalise them, would themselves be irrational. Now, we do not deny that there can be occasions on which such considerations do cause our beliefs, and we embrace the conclusion that when this occurs, the resultant belief is irrational. Nonetheless, there is now no reason to accept that these cases are the norm. Owens needs to say that pragmatic considerations operate in the unconscious mind to give the conscious impression of sufficient epistemic reason, to avoid the conclusion that all of our beliefs are irrational. However, we have removed this motivation for saying that this is the case: we have argued that Owens has not ruled out the idea that our beliefs are rationalised independently of such pragmatic considerations. So, why say that such considerations conspire at the unconscious level? We have no way of definitively demonstrating that they do *not* – we cannot appeal to the phenomenology of belief-regulation as *ex hypothesi* these considerations would operate outside of our

awareness. All we can show – as we have done – is that there is no need to think of the doxastic deliberator as unconsciously aware that they would have to epistemically rule out sceptical hypotheses if there were more time etc. Instead, why not think of the doxastic deliberator as aware that the only justification they need for their belief that there is a wine glass in front of them is that this appears to be the case?

4.7 Responsibility and Reflection

So far, then, we have seen no reason to deny that the control one has over one's beliefs is to be thought of on the model of reflective control discussed above. This picture, if correct, would capture our ordinary practices of holding one another responsible for our beliefs. In order to see this, it is worth thinking about these practices in a little more detail. Boyle (2012, pp. 4-6) effectively portrays a number of significant features:

"…if a person believes P, we expect him normally to be able to answer the question *why* he believes P, in the following sense: we expect him to be able to discuss what convinces him that P is true, what grounds he has for affirming this proposition. I do not mean that we expect a person always to be able to produce specific grounds for his beliefs: plainly, people can hold beliefs for which they do not have specific grounds. But the point that we should notice is that, even when a person admits to lacking grounds for a given belief, *he thereby accepts the presupposition of the question* – that he is in a position to speak for whatever grounds he has…

We normally address our questions and criticisms entirely to the soundness of the propositions… [a person] believes, criticizing them or the grounds he gives for them. And we do not merely make such criticisms *of* his beliefs; we address them *to* him: we ask *him* why he believes something so outlandish, why he accepts such a manifestly unreasonable argument, etc. If his belief on a certain point is demonstrably wrong or ill-grounded, we expect him to reconsider the matter – and we expect his reconsideration, not merely to change his self-assessment, but to change his first-order beliefs themselves. We thus seem to treat his believing as a circumstance that is in some sense up to him, one for which he is responsible in a specific way."

A number of connected aspects of our practices of holding one another responsible are being presented in the above passages. The most basic point is that our practices presuppose an extremely tight connection between possession of and rational responsibility for a belief: the why-question can always be asked with respect to a person's belief, and we expect them to be capable of providing some sort of response. We level this why-question at the believer themselves, and in so doing we treat them as occupying a position from which they may answer the question themselves, *speaking for* whatever grounds they may have. When the grounds the person presents are manifestly unsatisfactory, we treat the person themselves as rationally criticisable, and we expect that their acknowledgement of this criticism will bring with it a reappraisal of the belief that warranted criticism, which will often issue in revision or extinction of the offending belief.

Now, the reflective control model can accommodate all of these features. We suppose that the believer themselves is always in a position to speak for whatever grounds they may have, because we suppose that the believer is always capable of reflecting on their epistemic reasons, and such reflection will put the believer in a position to articulate their reasons, such as they are. Thus, to ask the why-question is to invite the believer to share the results of their reflection upon their reasons. We treat the why-question as always applicable to the believer because we suppose that the believer can always undertake such a process of reflection, whether they already have or not. It is for this reason as well that we treat the believer themselves as rationally criticisable if the grounds their refection discloses are judged to be lacking. In criticising a person's beliefs as irrational, for example, one is not simply making a disparaging remark about the belief itself, as one might sneer at a neighbour's unkempt lawn, but one is criticising the person themselves for ever having held such a belief. Even if the person has never reflected on their belief until the occasion of criticism, they could have, and they should have: had they done, they could have seen the inadequacy of their reasons for themselves, and could have abandoned their belief on this basis. Given that they did not, the present situation, in which they hold a criticisable belief, is in part of their own making, and this is why in criticising their belief we do more than simply bring to their attention the faultiness of their belief – we criticise them for ever having held it. Thus, our rational criticism of one another is fruitfully understood as presupposing reflective control: we criticise one another

for not having already exerted such control. This presupposition also explains our expectations about the results of such criticism. We can only bring into view the inadequacy of a person's ground by getting them to engage in precisely the process of reflection we criticise them for not having already undertaken. Given that they have gone through this process at the point at which they acknowledge rational criticism, we expect their acknowledgement of the criticism to be coeval with an exercise of reflective control, and thus we expect them to abandon or revise the offending belief.

In addition to drawing to our attention the presuppositions of our practices of holding one another responsible for our beliefs (which we have argued can be accommodated by the reflective control model), Boyle also highlights an unhelpful assumption that could be understood as underlying the reflective control model. According to this assumption, our control over our beliefs consists in our capacity to make a *judgement* by which we can *affect* what we believe. On this picture, as he puts it:

"…a person's agency can get no nearer to her beliefs than to touch them at their edges, so to speak" (forthcoming a, p. 3)

The decisive difficulty with this assumption is motivated by Boyle with the simple observation that judging that *p requires* believing that *p*. That is, one cannot judge that *p* unless one also believes that *p*. As Boyle (forthcoming a, p. 12) says:

"My conscious act of judging P must be expressive of my having settled on a view about whether P, namely: Yes, indeed, P. But it is hard to see how this can mean anything less than: it must be expressive of my believing that P."

Now, maintaining that this is the case is not to deny that sometimes the belief that *p begins* with the conscious judgement that *p*. It can be that one consciously judges that *p* at the first moment when one believes that *p*, whilst aware that one did not believe that *p* before that moment. But this is obviously distinct from saying that the judgement that *p caused* the belief that *p*.

Nonetheless, we need not think of reflective control as resting upon such an assumption.[28] Rather, we can and should maintain that in reflecting and judging that

---

[28] Boyle (forthcoming a, p. 9) does not think that it does: his "complaint is not that… [many authors] unambiguously endorse a wrong view, but that they do not unambiguously endorse a right one."

p one *thereby* alters one's beliefs. There is no further link in the causal chain. As we emphasised in the previous chapter, to be consciously satisfied that *p* simply *is* to consciously believe that *p* – and what more could such satisfaction amount to than a judgement that *p*?

In sum, then, we have argued that Owens has not demonstrated that we do not have reflective control over our beliefs. Furthermore, we have argued that our possession of this power of reflection would explain the manner in which we hold one another responsible to epistemic norms. Finally, we have noted that we ought not to think of reflection as involving the making of a judgement which then *affects* our belief – rather, consciously judging that *p is* consciously believing that *p*.

Conclusion

In this chapter, we have explored the nature of the correctness condition, which we postulated in order to explain the transparency of doxastic deliberation. We began by briefly exploring what it *means* to call a true belief a *correct* one. We noted that this claim can be understood variously, but, following Wedgwood (2002), we adopted a working understanding of it as commending a correct belief for being in some way *right*, or *appropriate*. We went on to suggest that we can derive two further epistemic norms from the correctness condition – what we called *normative exclusivity*, and *normative demandingness* – which inform us as to when a particular belief is *rational*. According to normative exclusivity, one ought to form the belief that *p* solely on the basis of epistemic reason in favour of *p*. According to normative demandingness, one ought to form the belief that *p* on the basis of sufficient epistemic reason in favour of *p*. We then considered two main objections to the idea that belief is subject to the epistemic norms we had arrived at. The first of these was the pragmatist's claim that what one should believe is often answerable to practical norms. We showed that the pragmatist's claim results from a confusion of two questions: *should I believe that p* and *should I try to get myself to believe that p*. Although the latter is answerable to practical norms, the former is answerable solely to epistemic norms. The second line of objection we explored to the idea that beliefs are answerable to the correctness condition and the derived norms of exclusivity and demandingness is the idea that we do not have the right kind of control over our beliefs to render this answerability coherent. We argued that Owens is wrong to

conclude that our beliefs are not under our reflective control. Given the theoretically unattractive consequence of his view – that we are not directly responsible for our beliefs in the way that we think we are – the burden of proof is firmly with Owens to establish his position beyond doubt. This he has not achieved, as his position rests upon a contentious assumption: that we need rule out all competing hypotheses in order to have sufficient epistemic reason in favour of *p* and, hence, rationally form the belief that *p*. Given that this is far from obviously correct, and given the theoretically unpalatable consequences of accepting it, it is open to us to maintain against Owens that pragmatic considerations do not ordinarily play any role in determining when we understand ourselves to have sufficient epistemic reason, and that, when they do, the resultant belief is irrational. Hence, there is no reason to reject the reflective control model – our reasons for rational belief are solely epistemic and so we can reflect upon them to marshal our beliefs. Furthermore, we argued that reflective control can capture the manner in which we hold one another directly responsible for our beliefs. And, finally, we noted that such reflective control must not be understood to involve making a judgement, which then affects one's belief from the outside. Rather, making a judgement *is* itself to believe.

Summary of the Position So Far

We will now move on, in chapters five and six, to explore two irrational phenomena – self-deception and delusion – which have more to teach us about the nature of belief. Before we do so, however, let us summarise what we have achieved so far. We began in chapter one by considering two difficult cases when it comes to belief attribution: Implicit Racist and Precipice. We noted that it seems that in order to explain the patterns in the behaviour of both Juliet and the woman in Precipice, we should appeal to the idea that there is some kind of significant relationship between belief and epistemic reasons. That is, in order to explain why all of their considered behaviour hangs together, as does all of their unconsidered behaviour, we should view the former as pertaining to their consideration of their reasons – and hence what they believe – and the latter to some other mental state. In chapter two, we set out to explore the contention that belief and reasons are importantly related. We noted that there are two key phenomena pertaining to belief, which appear to point in opposing directions. The first is what Shah has called the *transparency* of doxastic deliberation, the second, what we have called the *heterogeneity* of belief regulation.

Transparency seems to point to the idea that belief and reasons are importantly interrelated, and heterogeneity towards the conclusion that this is not the case. However, we argued that we can explain transparency, in spite of heterogeneity, in terms of Williams' claim that *beliefs aim at truth* if we understand this claim as follows. Regardless as to how beliefs are regulated outside of consciousness, it is constitutive of belief that if and when it is consciously regulated, it must be so-regulated solely for truth. This is what we have called the *regulation condition*. We then argued that the explanation as to why this regulation condition is constitutive of belief is that belief is at base a normative phenomenon: the fundamental constitutive fact about belief is that the belief that *p* is *correct* if and only if *p*. We called this claim the *correctness condition*. Given that correctness is constitutive of belief in this way, the most fundamental explanation of transparency runs as follows. Considering *whether to believe that p* is considering *whether one should believe that p*. This being the case, insofar as one understands the concept the belief, one understands that it is answerable solely to epistemic norms, the most basic of which is the correctness condition. So, given that one's question is *whether one should believe that p*, it would only be coherent to consider *whether p*. Having thus explained transparency, we went on to further explore the nature of both the regulation and correctness conditions in turn. Beginning with the former, we saw that regulation is not to be understood in terms of the doxastic deliberator themselves possessing a literal aim: rather, their belief aims at truth, simply insofar as coming to consciously believe that *p* is coming to be consciously satisfied that *p*. Then, turning to correctness, we saw that we can derive two subsequent epistemic norms from the correctness condition itself, which instruct us as to when a belief is rational: normative exclusivity and normative demandingness. Having thus defended and clarified the sense in which beliefs aim at truth (and hence, that there is an important relationship between belief and reasons), let us now progress to consider two irrational phenomena that promise to teach us more about the nature of belief.

5. Belief and Self-Deception

Our sense that there is an important relationship between belief and reasons was first prompted by our attention to two cases in which belief attribution is controversial. We then proceeded to develop, over the last three chapters, an account of the nature of belief designed to tie reasons and belief to one another at a constitutive level. With this account in hand, we will now return to consider some further cases in which belief attribution is problematic: cases involving self-deception and cases of delusion. We do so with the intention both of learning more about such cases *via* the verdicts our account demands, and more about the requirements of our account from its treatment of such cases. In the present chapter we consider self-deception, and in the final chapter, delusion.

An obvious place to begin when trying to understand self-deception is to look to ordinary cases of deception, in which one person deceives another – *other-deception*. In typical cases of other-deception, *A* believes that *p* and desires that *B* does not believe that *p*, but rather that they believe that *~p* instead. So, *A* intentionally brings it about that *B* believes that *~p*, whilst continuing to believe that *p* themselves. If we use this as a model for self-deception, we are obliged to think of the self-deceiver as believing that *p* (their undesired belief), as well as intentionally bringing about in themselves and sustaining the belief that *~p* (their desired belief). This model of self-deception spurs two philosophical problems: (i) the *intention problem*, which is to specify whether or not the self-deceiver possesses a self-deceptive intention; and (ii) the *doxastic problem*, which is to specify whether the self-deceiver holds either or both of the relevant beliefs. In this chapter, given our exclusive interest in belief attribution, we will be concerned to arrive at a solution to the doxastic problem. It has often been thought that a solution to this problem requires a solution to the intention problem, on the grounds that the chief obstacle to attributing the self-deceiver their desired belief is the idea that they intend to form that belief; thus, determining whether they possess the relevant intention will directly inform our verdict on which beliefs they hold. However, we will show in this chapter that a conclusive solution to the doxastic problem can in fact be arrived at independently of considerations of whether they possess a self-deceptive intention. Hence, our interest in the self-deceiver's beliefs will not bring with it an account of the status of their motivation.

In recent times, the strategy of attributing the self-deceiver both the belief that *p* and the belief that ~*p*, as invited by the other-deception model, has been thrown into doubt. On the one hand, there are theorists who claim that, in paradigm cases, the self-deceiver holds their undesired belief and not their desired belief (e.g. Robert Audi (1997), Eric Funkhouser (2005), Tamar Szabひ Gendler (2008a)), and on the other hand, there are those who claim that things are the other way around: in the paradigm case, the self-deceiver does not hold their undesired belief, but they do hold their desired belief (e.g. Annette Barnes (1997), Alfred Mele, as recently as (2010)). We will maintain that an option which all parties seem largely to overlook is in fact preferable to the alternatives discussed. We will argue for a position that we will call *doxastic minimalism*, according to which the self-deceiver neither believes that *p* nor believes that ~*p*. We will present both a negative and a positive argument for this conclusion. We will begin by arguing that no explanatory *need* to attribute the self-deceiver either belief has been demonstrated. As we saw in our first chapter, folk psychology contains a rich array of explanatory tools – we do not need to rely solely on belief and desire – and, if we bear this firmly in mind, we will see that all of the phenomena that have been adduced to point to the idea that the self-deceiver holds either one belief or the other can be explained without recourse to belief. We will then turn to the understanding of belief we have been developing in this thesis to reveal that not only is it unnecessary to attribute the self-deceiver either belief, we *should* not, either. We will argue that in the paradigm case of self-deception, the self-deceiver cannot be said to hold either their undesired or desired belief. Their undesired belief would need to be unconscious in the strong sense that they would be unable to reflect upon it. However, given the intimate connection between reflective control and possession of belief argued for in the previous chapter, this is as much as to say that the self-deceiver's putative belief would fail to satisfy the correctness condition, but a state that fails to satisfy this condition cannot be a belief. In contrast, the self-deceiver does reflect upon reasons in favour of their desired belief that ~*p*. However, when they do so, the unease that they feel surrounding the topic of *p* precludes us from thinking of this reflection as regulation of a belief solely in accordance with truth. Thus, our account of belief teaches us that the paradigm case of self-deception does not involve belief, and self-deception teaches us that our account militates against beliefs that are unconscious in the present strong sense, and requires that the satisfaction attendant upon viewing one's epistemic reasons as

sufficient should be incompatible with the feelings of doubt and unease exhibited in self-deception.

We will proceed as follows. We will begin with a rough-and-ready characterisation of what we will mean by a *paradigm* case of self-deception. We will then outline the four main options one has with respect to attributing the self-deceiver beliefs. We will go on to take each purported belief – the self-deceiver's undesired belief and their desired belief – in turn and consider the explanatory role it is alleged to play in isolation from the other. We will argue that, considered independently, there is no explanatory need to postulate either belief. Of course, our negative conclusion does not follow straightforwardly from this, as there may be some reason(s) to think that the self-deceiver must hold at least one of the beliefs in question. Thus, we will go on to examine how attributing the self-deceiver one belief might be supposed to interact with failing to attribute them the other, as well as some additional reasons one might think that the self-deceiver need hold at least one of the beliefs. We will argue that none of these considerations are convincing. So, we will conclude that, in paradigm cases of self-deception, no explanatory need to attribute the self-deceiver either their undesired or their desired belief has been demonstrated. We will then present positive reason to resist attributing the self-deceiver either the belief that *p* or the belief that ~*p*, by appeal to the requirements of the regulation and correctness conditions on belief argued for in the preceding chapters. In so doing, we will learn more about the requirements of the account that we have developed.

5.1 Paradigm Cases

Self-deception is attributed to people in a wide variety of cases. In fact, there is so much disparity in cases of self-deception that a unified account of the phenomenon may be difficult to provide. Furthermore, there is disagreement even about how to account for paradigm cases. So, in what follows, we will limit our enquiry to paradigm cases of the phenomenon: unless specified otherwise, when we refer to self-deception, we refer to paradigm cases of the phenomenon.

So, what counts as a paradigm case of self-deception? It seems that in some situations, the self-deceiver attempts to acquire a belief that they *desire,* and in other cases a belief that they do *not* desire. Consider the contrast between the husband who desires to believe that his unfaithful wife is not having an affair and so deceives

himself that this is the case, and the jealous husband who deceives himself that his faithful wife is having an affair, although he does not desire to believe this. We will include only the former kind of case, in which the self-deceiver desires to believe that ~p, under the banner of *paradigm* cases of self-deception with respect to p. The kinds of cases we will be concerned with are described by Audi (1982, p. 134) here:

"…the lover who cannot bear the thought that his beloved is drawing away, the alcoholic who cannot admit that he is unable by himself to stop drinking, the terminal patient unable to face his death, and the athlete unable to reconcile himself to his waning powers."

5.2 The Options

The four main options for characterising the self-deceiver's beliefs are as follows:

(1)   The self-deceiver both believes that *p* (their undesired belief) and believes that ~p (their desired belief). (Donald Davidson (e.g. 1986) is the most famous proponent of this position.)

(2)   The self-deceiver believes that *p* (their undesired belief) but they do not believe that ~p (their desired belief). (Proponents include Audi (1997), Eric Funkhouser (2005), Tamar SzabU Gendler (2008a).)

(3)   The self-deceiver does not believe that *p* (their undesired belief), but they do believe that ~p (their desired belief). (Proponents include Barnes (1997), and Mele, as recently as (2010).)

(4)   The self-deceiver neither believes that *p* (their undesired belief) nor believes that ~p (their desired belief).

So, if we take the case of the husband who desires to believe that his unfaithful wife is not having an affair and becomes self-deceived over her infidelity, according to option (1), he both believes that she is having an affair and that it is not the case that she is having an affair. According to option (2), he believes that she is having an affair but he fails to bring himself to believe that it is not the case that she is having an affair. According to option (3), he does not believe that she is having an affair, but rather manages to bring himself to believe that it is not the case that she is having an

affair. And, finally, according to option (4), he neither believes that she is having an affair nor believes that it is not the case that she is having an affair.

The major difficulty with (1) is to explain how it is that the self-deceiver can both believe that p and believe that $\sim p$, and yet fail to put these two beliefs together to form the impossible belief that p and $\sim p$.[29] Davidson (1986) attempts to explain how this can be so by appealing to the idea that the mind is divided. However, it is far from clear that his attempt to do so is successful, as it stands.[30] Given that this is the case, all parties to the debate we are about to enter assume for the sake of argument that the self-deceiver's beliefs are not to be conceived as *per* (1). We will follow suit for now, but, of course, by the end of this chapter we will have a positive argument in favour of (4), and hence we will have provided reason for denying (1).

We will now begin to make our case for (4) by presenting our negative argument: we will argue that there is no explanatory need to attribute the self-deceiver either belief. We will start by taking each of the self-deceiver's purported beliefs in turn and examining the arguments for attributing it to them, when considered in isolation from the other belief. First: their undesired belief.

5.3 The Undesired Belief: A Negative Case

We will begin by considering the idea that the self-deceiver's *behaviour* reveals them to hold their undesired belief. Audi (1982) (a proponent of (2)) argues that this is the case: the self-deceiver's behaviour reveals them to unconsciously hold their undesired belief. Before we progress, it is worth pausing briefly to consider what is meant by the claim that they hold this belief *unconsciously*. In particular, it is important to distinguish between the indisputable fact that there can be unconscious beliefs that one merely *happens* not to be conscious of at a certain time, and the claim that there can be unconscious beliefs that there is some psychological barrier to consciously countenancing. Audi (1982, p. 137) conceives of the self-deceiver's belief along the latter lines:

---

[29] This belief is impossible, provided the believer does not also believe that there can be true contradictions. Graham Priest (2006) is an example of someone who holds that some contradictions are true, but his view is a minority one and we will set it aside.

[30] For a convincing criticism of Davidson's position, see David Pears (1984).

"…unconscious beliefs are very much like conscious ones, apart from two major differences: (i) if they manifest themselves is *S*'s consciousness, he is very unlikely, without special self-scrutiny or prompting from someone else, to attribute these manifestations to them; and (ii), he is, with the same exceptions, very unlikely to explain actions of his which are due to them, *as* due to them. But – and this is the important point here – they do tend to manifest themselves in consciousness and behavior, and in essentially the same way as conscious beliefs, though usually less frequently…

Roughly, we might say that *S*'s belief that *p* is unconscious if and only if (1) he does not know or believe he has it, and (2) he cannot come to know or believe he has it without either outside help…or some special self-scrutiny."

We will now consider two of Audi's examples of paradigm cases of self-deception that he claims testify to the fact that the self-deceiver unconsciously holds their undesired belief. Against Audi, we will argue that we do not need to resort to the claim that the self-deceiver holds their undesired belief in such cases in order to explain those aspects of their behaviour that superficially point to this conclusion.

First, there is Audi's case of Ann. Audi (1982, p. 134) asks us to imagine that:

"… Ann is dying of cancer and is aware of many facts, such as her long, steady decline, pointing to this outcome, though no one has told her that her case is terminal and she has avoided letting her doctor give her a prognosis. Suppose further that she talks of recovery and discusses her various plans for the long future…the facts pointing to her death are not unmistakably prominent and her talk of recovery is apparently sincere…she has better than average medical knowledge…[but] (among other things)…her talk of recovery lacks full conviction (or exhibits too much apparent conviction), and…is often followed by depression or anxiety."

He describes how Ann "asks about funeral arrangements "just in case"" and "rewrites her will" etc. (Audi, 1982, p. 139). Such behaviour warrants the attribution of the unconscious belief that she will soon die of cancer to Ann, according to Audi. But it seems that there are many other ways of accounting for this behaviour on Ann's part that do not involve attributing her an unconscious belief. Mele (2009, p. 273) (a proponent of (3)) suggests that we say that she consciously believes "that

there is a significant chance" that she will soon die of the disease, which, together with her relevant desires, explains why she makes out her will etc. However, this explanation seems untrue to the phenomenology of self-deception. It seems that part of the experience of being self-deceived is the avoidance of confronting the possibility of the undesired situation too explicitly, so to attribute the self-deceiver the conscious belief that there is a chance that this situation obtains rings untrue. Rather, in Ann's case, a ready folk psychological explanation is that Ann is *anxious* about the state of her health and *suspects* that things are worse than she is currently aware, and that she may find out that something she *fears* is the case by probing the matter too much with her doctors etc., so she avoids confronting the matter in the manner in which she would need to in order to form the true belief that she is terminally ill, or even that there is a significant chance that this is the case. Nonetheless, her anxiety and suspicions are enough to prompt the actions on Ann's part that seem to suggest that she believes her undesired belief.

Mele (e.g. 1997, p. 96) also thinks that a suspicion that *p* can often carry the majority of the burden of explaining the self-deceiver's behaviour that appears to point to their holding their undesired belief. However, one might object that to suspect that *p* is the case involves being disposed, *ceteris paribus*, to investigate in an even-handed manner whether or not *p* is in fact the case. But, rather than straightforwardly investigating whether or not their undesired state of affairs obtains, self-deceivers typically do all they can to avoid discovering that it does. Does this suggest that a suspicion is not the appropriate mental state to appeal to here? Well, for a start, it is unclear that suspecting that *p* does disposes one – *ceteris paribus* – to behave in such a way. Furthermore, even if it did, in the case of the self-deceiver, all else is not equal. The self-deceiver is *anxious* that *p* and they *desire* not to believe that *p* etc. and this explains why their suspicion that *p* does not prompt them genuinely to seek out evidence as to whether or not *p* is the case, but instead disposes them to misinterpret and avoid evidence that *p* and seek only evidence that ~*p*. So far, of course, we are yet to provide a positive argument for regarding a suspicion-based explanation as superior to a belief-based explanation of the self-deceiver's behaviour. We will consider whether such an argument can be provided after our negative case is complete.

The second case of Audi's (1972, p. 48) we will examine is that of a woman who is self-deceived about the fact that her husband is a "failure". Again, Mele suggests that a belief other than the unconscious belief that her husband is a failure can explain the woman's behaviour that appears to point to her holding this belief. He thinks that the belief that "her husband has failed in many ways" could help explain the woman's negative behaviour towards her husband, or her treating him as she would a failure, without her believing him to *be* a failure (Mele, 1982, p. 161).

Again, it seems that the phenomenology of self-deception precludes us from thinking of the woman as too readily attending to any such belief. However, Mele's suggestion can be taken on board, provided that we recognise that folk psychology provides the nuanced language necessary to describe the case accurately: a ready explanation of the woman's negative behaviour towards her husband involves her possessing a *dim awareness* of her husband's many failures, combined with a *resentful attitude* towards these failures. Once more, there is no explanatory need to suppose that the self-deceiver unconsciously holds her undesired belief.

Other than the self-deceiver's behaviour, are there any further considerations that appear to support the attribution to them of their undesired belief? A second consideration appeals to the idea that their holding such a belief has a *causal* role to play: it helps cause and sustain their self-deception. Let us take the case of the husband who is self-deceived over his wife's infidelity as an example. It might be thought that it is the belief that his wife is having an affair (along with his strong desire not to believe this, for example) that prompts him to enter into and sustains his self-deception. However, it seems that there is no need to attribute the husband his undesired belief in order to explain the initiation and sustenance of his self-deception. A suspicion that his wife is unfaithful to him on the part of the husband, combined with a strong desire to believe that this is not the case, fill this explanatory role.

A third consideration is the self-deceiver's *phenomenology*, to which we have already briefly alluded. William Talbott (1995) considers why this might be thought to point towards the self-deceiver's holding their undesired belief. He claims that it is the:

"…almost palpable quality of emotional resistance to doubts about… [their desired belief] or evidence against… [it] that, I believe, most inclines us to attribute to the self-deceived person at some level a recognition that… [their undesired belief is true]. If she does not really recognize that… [her undesired belief is true], why does the evidence that…[it is] produce such an emotionally charged reaction?" (Talbott, 1995, pp. 44-45)

However, as Talbott says, it seems that such reactions can be explained simply by the self-deceiver's awareness that the evidence *could* be taken by them to point towards their undesired belief, and that if it were, the consequences would be disastrous for them. A suspicion that *p* may even seem to be present in some cases, but, again, it is not explanatorily necessary to attribute them their undesired belief.

A fourth consideration we must consider is what we might call the self-deceiver's *inputs*. Consider the balding man who is self-deceived about his steadily receding hair-line (this kind of example is employed by Davidson (1986)). We might ask: how can he fail to believe that he is balding? After all, he owns a mirror. However, this line of thought presumes that the biasing strategies employed by the self-deceiver, which both control what the self-deceiver's inputs are in the first place, and distort the information once it arrives, cannot be substantially successful. Clearly, the uneasiness the balding man feels when his wife motions as to ruffle his hair, and the manner in which he avoids dwelling for longer than a split second on his image in the mirror in the mornings suggest that the biasing strategies are not wholly successful: enough information has gotten through to make him extremely uncomfortable around the topic of his hair; nonetheless, the suggestion is that the biasing strategies employed by the self-deceiver enable him to avoid forming the belief that he is balding. To stipulate that this cannot be so is to beg the question against one who claims that these strategies *can* be substantially successful.

Fifthly, we might draw attention to what ex-self-deceivers typically *say* of their experience of self-deception having come out of their self-deceptive state. When they are no longer self-deceived, they may say things such as *I knew all along/deep down that p was the case*, for example. However, given that the suggestion is that they believed that *p* unconsciously, in Audi's sense, it is not obvious that the self-deceiver stands in a privileged relation to their prior self-deceptive state. If they say such

things, they would presumably be inferring that these things are the case, just as someone else may infer that the self-deceiver believed their undesired belief, based upon their behaviour. (Barnes (1997, p. 89) makes a similar point with respect to the self-deceiver's purported intention.)

A sixth consideration is that we often hold self-deceivers responsible for their self-deception – be it rationally, or morally, or sometimes both. Perhaps this would be unwarranted if they did not hold their undesired belief. However, even granting that we are on occasion correct to censure self-deceivers, this can be explained with or without the idea that they believe their undesired belief. Perhaps they are to be criticised simply for engaging in their biased evidence gathering, for example.

So far, then, we have considered six apparent motivations for attributing the self-deceiver their undesired belief: their behaviour, the causation and sustenance of self-deception, the phenomenology of self-deception, their inputs, what people say of their own past mental states when they are no longer self-deceived, and the fact that we often hold self-deceivers to account over their self-deception. We have argued that none of these considerations demonstrate an explanatory need to suppose that the self-deceiver holds their undesired belief. In particular, we have drawn attention to the fact that folk psychology is sufficiently rich to be able to explain the behaviour and phenomenology of self-deception without appealing to the idea that the self-deceiver unconsciously believes their undesired belief. By way of example, we have suggested that a *suspicion* that *p*, combined with *anxiety* that *p*, and a *desire* not to believe that *p*, but to believe that ~*p*, may meet the explanatory charge in some cases.

5.4 The Desired Belief: A Negative Case

We will now turn to the self-deceiver's desired belief. Again, we will begin by considering whether or not their behaviour warrants the attribution of this belief to the self-deceiver. As with their undesired belief, *prima facie* the self-deceiver behaves as if they believe that ~*p*. Audi's Ann, who is dying of cancer, makes plans for the long-term future and talks as if it is not the case that she will soon die etc. However, Audi appeals to his example of Ann in order to argue that the self-deceiver's behaviour does *not* support the attribution of their desired belief to them. He describes how Ann's "talk of recovery lacks full conviction (or exhibits too much apparent conviction)" and so does not justify the attribution to her of the belief that it

124

is not the case that she will soon die of cancer (Audi, 1982, p. 134). Furthermore, according to Audi (1982, p. 139), behaviour that seems to evidence the self-deceiver's holding their desired belief could equally be explained in terms of their "wants, needs, etc. that explain why the unconscious belief is unconscious in the first place." This is somewhat suggestive. Is there more that could be said here?

We could begin by looking to Audi's own suggestion that the self-deceiver *sincerely avows* their desired belief that ~*p*, where sincere avowal is understood as not entailing belief. Audi thinks that the self-deceiver sincerely avows that ~*p* insofar as they lack the belief that they are speaking falsely in so doing, lack any intention to deceive, are not disposed to tell themselves or those they trust that *p*, and so on. Is the sincerity of one's speech acts secured simply by failing to believe that one is speaking falsely? It seems that Audi is correct that the person who speaks without any belief that what they are saying is false etc. is not straightforwardly *insincere*. However, it also seems that more is required of one who is sincere in what one says: in order to speak sincerely, perhaps one need believe what one says, or perhaps one need only believe that one believes it.

We will not pursue this issue, however, as it is not clear that we need think of the self-deceiver's speech acts as sincere in any case. It seems truer to the phenomenology of self-deception to say that in asserting their desired belief that ~*p*, the self-deceiver experiences a certain amount of conflict – a niggling feeling etc. As Audi (1982, p. 134) says, their speech acts often evidence this in lacking "full conviction (or exhibit[ing]… too much apparent conviction)". Given that this is so, one ready explanation of what the self-deceiver is doing when they assert that ~*p*, for example, is what Mele (1987a: 144) calls "acting as-if". As Mele explains, acting as if one's desired belief were true can help generate evidence in favour of one's desired belief, as one can be swayed by one's own behaviour. It can also "generate "social" evidence for a favored hypothesis": if the self-deceiver acts as if something is the case, then others may well respond to them as if this is the case, which generates further reassurance from the self-deceiver's point of view (Mele, 1987a: 158). As Mele (1987a: 157) says, "it is not difficult to see why someone may be motivated to act as if… [~*p*] by a desire… [to believe that ~*p*]."

We can also think of the self-deceiver as engaging in a kind of *internal* acting as-if, in addition to their external behaviour of this kind. Not only do they tell others that ~p, they continually *tell themselves* this as well. In addition to their *anxiety* that p, their *hope* that ~p, their *desire* to believe that ~p etc., such internal acting as-if could also help lead the self-deceiver to behave as if ~p.

So, there are viable alternatives for explaining those aspects of the self-deceiver's behaviour that appear to point to their holding their desired belief. Given this, let us move on to consider any further reasons one might have for attributing the self-deceiver their desired belief. Let us briefly consider the phenomenology of the self-deceiver once more: any attraction the self-deceiver feels when the thought ~p is entertained is readily explicable in terms of their desire to believe that ~p and to avoid believing that p. Also, as before, their subsequent assertions with respect to their past self-deception need be granted no special authority. If an ex-self-deceiver were to say *I really managed to convince myself that ~p*, the appropriate response would seem to be to doubt their apparent memory of their past state. Paradigm cases of self-deception necessarily involve feelings of conflict which would preclude such a sense of conviction. And fourthly, any responsibility we are likely to attribute to the self-deceiver for their self-deceptive state, even if well-placed, does not require the idea that they believed the false, desired belief. They are rationally (and perhaps sometimes morally) criticisable for their biased evidence gathering.

In this section, we have considered four apparent motivations for attributing the self-deceiver their desired belief: this time, we have considered their behaviour, the phenomenology of self-deception, what people say of their own past mental states when they are no longer self-deceived, and the fact that we often hold self-deceivers to account over their self-deception. As with the self-deceiver's undesired belief, we have argued that these considerations demonstrate no explanatory need to attribute the self-deceiver their desired belief.

5.5 Putting the Two Together: The Negative Case Continued

So far, then, we have seen that, considered in isolation, no explanatory need has been demonstrated to attribute the self-deceiver either their undesired belief that p, or their desired belief that ~p. Of course, it does not follow straightaway from this that in the paradigm case there is no explanatory need to attribute the self-deceiver either belief.

What if there were some reason(s) to think that they must have at least one of the beliefs in question? We will begin by considering the possibility that attributing the self-deceiver one of their beliefs is necessary for withholding the attribution of the other belief.

According to position (2), on which the self-deceiver holds their undesired belief that *p*, but not their desired belief that ~*p*, it could be thought that their undesired belief helps to explain why the self-deceiver fails to attain their desired belief. Given the assumption that they do not hold contradictory beliefs, the idea could be that the self-deceiver is frustrated in their attempts to acquire their desired belief by their prior possession of their undesired belief.

On position (3), according to which the self-deceiver does not hold their undesired belief, but does hold their desired belief, the fact that we attribute the self-deceiver their desired belief could be understood as explaining why we fail to attribute them their undesired belief. Again, given the assumption that they do not hold contradictory beliefs, the fact that we are prepared to attribute them their desired belief might be thought to explain why we cannot attribute them their undesired belief.

However, it seems that these explanatory roles can be met without attributing the self-deceiver either belief. Let us take the case of the husband who is self-deceived over his wife's infidelity once more. According to the doxastic minimalist, the husband begins to *suspect* that his wife may be having an affair and, immediately, his defences go up. He strongly *desires* both that his wife is faithful and to believe that this is the case. The combination of this desire and his suspicion explains why he begins to avoid information in favour of his feared conclusion and seek out evidence against it, and distort any evidence he finds for his own purposes. Engaging in such biased evidence gathering, he prevents himself from coming to hold his undesired belief that his wife is having an affair. Even so, there is significant disquiet in his mind regarding the issue: he is *anxious* that it not be the case that his wife is having an affair, he has niggling *doubts* about the information he already has, but he *hopes* that it is not the case that she is unfaithful to him. Nonetheless, his niggling doubts and suspicions prevent him from attaining the belief that it is not the case that his wife is having an affair.

On this construal of the case, the explanation as to why the husband fails to attain his desired belief is that his niggling doubts and suspicions to the contrary prevent him from so doing. We need not stipulate that only a belief that p can prevent one from forming the belief that $\sim p$. Similarly, there appears to be no reason to insist that only the possession of his desired belief prevents the husband from failing to hold his undesired belief. The biasing mechanisms described ensure that he fails to hold his undesired belief, whether or not they result in his coming to believe his desired belief.

So, we have seen that attributing the self-deceiver one of the beliefs in question is not necessary for withholding the attribution of the other. But are there any further reasons to think that the self-deceiver must hold at least one of the beliefs under discussion?

First, one might think that if someone feels drawn to both the belief that p and the belief that $\sim p$, yet holds neither, they are in a state of simple ambivalence with respect to p; they are not self-deceived about p. So, one might think, in order to distinguish self-deception from ambivalence, we must attribute the self-deceiver either the belief that p or the belief that $\sim p$. However, it seems that there are at least two ways in which the doxastic minimalist can distinguish self-deception from simple ambivalence. First, we can draw attention to the fact that self-deception is necessarily motivated: the self-deceiver's desire to avoid their undesired belief and promote their desired belief, as well as their anxieties surrounding the subject-matter about which they are self-deceived, mark out self-deception from simple ambivalence. Secondly, we might point to the fact that if one is simply ambivalent about whether p, one will typically behave very differently than if one is self-deceived concerning p. For example, if one is merely ambivalent about p, *ceteris paribus*, one will affirm that this is so, whereas in the paradigm case, the self-deceiver will vehemently affirm that $\sim p$.

Secondly, one might think that in order to explain why self-deception was ever so called in the first place, one needs to retain something of the model of other-deception. Why is self-deception called *deception* if the self is neither the deceiver who believes the truth, nor the deceived who believes a falsehood? But the doxastic minimalist can maintain that self-deception acquired the name it did because the self-

deceiver's behaviour superficially suggests that the self-deceiver holds both their undesired and their desired beliefs (and, of course, because self-deception is apparently intentional), and so appears to be both deceiver and deceived. The doxastic minimalist does not need to appeal to the self-deceiver's possession of at least one of the beliefs in question to explain why the name given to the phenomenon under discussion is *self-deception*.

Finally, one might think that we need to attribute the self-deceiver either their undesired or their desired belief in order to explain the irrationality involved in self-deception. Perhaps the self-deceiver is irrational insofar as they believe the truth, but act as if it were not so, or because they have come to hold an irrational belief in a falsehood. However, the doxastic minimalist can readily explain the irrationality involved in self-deception without appealing to either belief: self-deception is (epistemically) irrational simply insofar as it involves the desire to avoid believing a certain proposition and to believe another, irrespective of their truth-values.[31]

5.6 The Positive Case

We have now completed our negative case in favour of doxastic minimalism. That is, we have argued that, as yet, no explanatory need has been demonstrated to attribute the self-deceiver either their undesired or their desired belief. We will now progress to make our positive case for the idea that we ought not to attribute the self-deceiver either the belief that p or the belief that $\sim p$. In so doing, we will tease out further consequences of the account of belief we have been developing.

Let us begin with the self-deceiver's undesired belief. The claim we are contending with is that the self-deceiver holds their undesired belief unconsciously, in Audi's sense. So far, we have seen that there is no explanatory need to say that this is the case, because attributing the self-deceiver the suspicion that p in addition to anxiety that p will also meet the explanatory need. It is now time to break the tie between these alternative accounts.

We have seen that it is constitutive of belief that *if* it consciously regulated, it is so-regulated solely in accordance with truth. However, the hypothesis we are

---

[31] This is not to say that undertaking a self-deceptive strategy cannot sometimes be practically rational, as we saw when we discussed pragmatism.

considering is that the self-deceiver unconsciously holds their undesired belief and hence is prevented from consciously regulating it. So, it seems that (R) will be of no help to us. However, we have also seen that it is (C) – the claim that it is constitutive of belief that it is correct if and only if its contents are true – that ultimately explains (R). What happens if we look directly to (C) for help? Well, in the previous chapter we saw that we can derive both normative exclusivity and demandingness from the correctness condition; consequently, both of these epistemic norms are also constitutive of belief. Furthermore, we argued that if we are to think of these derivative epistemic norms as constitutive of our belief then we must think of ourselves as having reflective control over our beliefs. That is, the manner in which we hold one another responsible for our beliefs reveals that we take it to be a constant possibility that one can reflect upon one's beliefs and thereby regulate them. However, it now becomes clear that a consequence of this claim about reflective control is that it renders the kind of unconscious belief Audi describes not merely unusual, but impossible. It is incompatible with our folk psychological sense that we are to be held directly responsible for our beliefs at all times that we can hold a belief without having the capacity to consciously reflect upon it at any given time. It is therefore a requirement of our account not just that a belief must behave in a certain way when consciously regulated (as described in (R)), but also that it must be constantly amenable to conscious regulation. Thus, we have arrived at a positive reason to prefer the hypothesis that the self-deceiver suspects that *p* as opposed to the hypothesis that they unconsciously (in Audi's sense) believe that *p*: such unconscious belief is impossible.

This brings us to the self-deceiver's desired belief. Those who suggest that the self-deceiver holds their desired belief of course do not maintain that it is prevented from conscious regulation. Indeed, paradigmatically, the self-deceiver *will* consider the proposition that *p*, and they will quickly dismiss it along the following lines: *Of course she's not having an affair. Our relationship is better than ever!* But isn't this simply conscious regulation for truth? Of course, the self-deceiver's subjective epistemic reasons are not objective epistemic reasons, but nonetheless, aren't they regulating their *belief* in line with the truth as they see it? It seems not. As has been admitted by all sides, in the paradigm case of self-deception, the self-deceiver feels a characteristic *unease* whenever the proposition *p* is entertained. They do not want to

dwell upon its truth-value for long, and we have suggested that this is because they suspect that all is not well. Now, given that this is the case, it seems that we ought not to allow that they are consciously regulating their belief that ~*p* for truth. They are uneasy about *p*, and this sense of unease – however faint – precludes the satisfaction that we have understood to be involved in conscious belief regulation. It is precisely this sense of unease that makes it more apt to describe the self-deceiver as acting as-if ~*p*, as opposed to consciously regulating the belief that ~*p* for truth when they think thoughts such as: *Of course she's not having an affair. Our relationship is better than ever!*

In sum, then, we have now argued that, not only is there no explanatory *need* to attribute the self-deceiver either the belief that *p* or the belief that ~*p*, but there is also positive reason not to do so. Furthermore, in articulating our positive case for doxastic minimalism, we have learnt more about our understanding of belief. With respect to the self-deceiver's undesired belief, we have argued that they cannot be said to hold this unconsciously in Audi's sense, because such unconscious belief is incompatible with the manner in which we hold one another responsible for our beliefs, which we explored in the previous chapter. With respect to the self-deceiver's desired belief, we saw that they cannot be understood to hold this, as the manner in which they regulate their representation that *p* is not compatible with (R). That is, we have learnt that the satisfaction involved in consciously believing precludes the sense of unease about *p* characteristic of self-deception.

Conclusion

In this chapter, we have explored the phenomenon of self-deception with the hope that it will reveal more to us about the nature of belief. We have argued for a position that we have called *doxastic minimalism* according to which, in the paradigm case of self-deception, the self-deceiver neither holds their undesired belief that *p* nor their desired belief that ~*p*. In so doing, we upheld our allegiance with the idea mooted in chapter one, that ordinary, folk psychology affords a rich array of potentially explanatory concepts, which stretch beyond the two old favourites – belief and desire. We proposed that we consider the self-deceiver's hopes, suspicions, doubts, and anxieties, to name a few, as opposed to thinking of the majority of their behaviour as revelatory solely of belief-desire pairs.

We began by presenting a negative case in support of this claim. We took each of the self-deceiver's purported beliefs in turn and argued that, considered on its own, there is no explanatory need to attribute it to them. We then argued that there is no need to say that, in the paradigm case, the self-deceiver need hold at least one of the beliefs in question. We began by showing that the attribution of one belief is not necessary to explain why we withhold the attribution of the other. The self-deceiver fails to attain their desired belief because of their niggling doubts and suspicions otherwise. They fail to attain their undesired belief because of the success of the biasing strategies they begin to employ as soon as they suspect that things are not how they would like them to be. Furthermore, we argued that the distinction between self-deception and simple ambivalence is easy for the doxastic minimalist to draw – either in terms of the self-deceiver's motivation, or their behaviour. They can explain why self-deception was so-called in terms of the surface appearances of the phenomenon. And, finally, the doxastic minimalist can account for the irrationality of self-deception without appealing to the self-deceiver's beliefs, by pointing to the fact that the self-deceiver is motivated to avoid one belief and attain another, irrespective of their truth-values.

After presenting this negative case in favour of doxastic minimalism, we mounted a positive case in its support and, in so doing, we discovered more about the conception of belief we have been developing throughout the course of this thesis. We argued that the self-deceiver cannot be said to hold their undesired belief unconsciously in Audi's sense, because the existence of such unconscious belief is incompatible with the claim that the correctness condition is constitutive of belief. As we argued in the previous chapter, we hold one another rationally responsible for our beliefs in a direct manner, which implies that we could – at any moment – reflect upon them. This precludes the existence of unconscious beliefs as Audi understands them. As for the self-deceiver's desired belief, we argued that they cannot be said to hold this, as the manner in which they consciously regulate the thought that p cannot be counted as solely for truth, as it includes a certain amount of unease on their part surrounding p.

6. Belief and Delusion

We now turn to a less pervasive and often more disturbing phenomenon than self-deception: delusion. Once more, we will consider the phenomenon in the light of the account of belief that we have been developing, with the hope of learning more both about delusion itself and the implications of our account of belief. The question this chapter will be organised around is: Does someone who is deluded that *p* believe that *p*? Those who answer this question in the affirmative we will refer to as *doxasticists* about delusion (e.g. José Bermúdez (2001), Lisa Bortolotti (2010)), and those who answer in the negative, *non-doxasticists* (e.g. Gregory Currie (2000), John Campbell (2001), Andy Egan (2008)). Now, in the previous chapter we observed that self-deception is a diverse phenomenon and hence we proceeded by discussing what we called paradigm cases. In this chapter we will proceed slightly differently. Given that one case of delusion can vary vastly from the next, our approach will be to consider typical features of delusion that will be possessed to a greater or lesser extent in any given case, and explore the manner in which such features will impact upon belief attribution. We shall not be concerned to legislate one way or the other as to whether all delusions do or do not involve the deluded individual believing the content of their delusion. Rather, we will argue that although some cases of delusion may best be understood as extreme cases of irrational belief, in other cases the delusion falls of the edge of the scale of irrational beliefs and no longer merits belief attribution.

Before we begin, it will be useful to re-state what counts, on our picture of belief, as an irrational belief. With this clarification in hand, we will be better able to adjudicate the question as to whether the deluded subject is simply in the grips of an extremely irrational belief, or whether they should be described as having fallen off the scale from the rational to the irrational, such that they no longer hold a belief of either kind. According to the analysis of the correctness condition developed in chapter four, the following prescription tells a subject when it would be rational to believe that *p*:

(Ra): One ought to believe that *p* solely on the basis of sufficient objective epistemic reason that *p*[32]

---

[32] For the remainder of this chapter, we will be concerned solely with epistemic reasons. Accordingly, we will now drop the qualifier *epistemic*.

Any belief that fails to comply with this prescription is to be counted as irrational. Now, we know from chapters two and three that it is impossible for a subject to consciously regulate a belief without thinking of themselves as complying with this condition, since we know that:

(R): It is constitutive of belief that, if it is consciously regulated, it is so-regulated solely for truth.

That is to say, it is not possible to hold the belief that *p* on the basis of conscious regulation without being satisfied that one holds the belief solely on the basis of sufficient objective reason that *p*. Nonetheless, there is a difference between being satisfied that one has met (Ra), and actually meeting it, and it is in the space held open by this difference that irrational beliefs may creep in. This difference is effectively expressed in the terminology of subjective and objective reasons: one has a subjective reason when one takes it that something is an objective reason for something else. (R) tells us that one may consciously form a belief only when one has sufficient subjective reason – that is, one takes there to be sufficient objective reason in support of believing *p* for one's belief to satisfy (Ra). In the good case, sufficient subjective and sufficient objective reasons will align, but a belief will be irrational and hence fail to meet (Ra) when one has sufficient subjective reasons in spite of having failed to latch on to sufficient objective reasons for believing that *p*. Now, with this in hand, the question as to whether the subject who is deluded about *p* believes that *p* can be re-expressed as the question as to whether the deluded subject holds fast to *p* on the basis of sufficient subjective reasons that *p*: that is, does the deluded subject hold fast to *p* because they are satisfied that they have discovered sufficient objective reason in support of *p*'s truth?

The considerations that we will present in this chapter are fruitfully understood as responses to this question. Having introduced delusional phenomena with reference to a number of psychological case studies, we will consider the claim that the content of a delusion is often not the sort of content that could be affirmed on the basis of sufficient subjective reasons, because it is incoherent, and one cannot take oneself to have sufficient subjective reasons for believing an incoherent content. We will argue that these considerations do little to support the non-doxasticist's case, because many of the delusional contents that are adduced as *prima facie* incoherent can in fact be

easily construed as coherent, once we recognise, with Tim Bayne and Elisabeth Pacherie (2005, p. 182), that "[t]he semantics of a word may reflect the fact that it is part of several scientific or folk-psychological domains". We will then turn to a thought associated with Willard Quine (e.g. 1960) and Davidson (e.g. 1973): namely, that treating *p* as subject to some tolerable degree of *rational integration* with one's other beliefs is constitutively involved in taking there to be sufficient objective reasons in support of *p* for rational belief that *p* to be possible.[33] In extreme cases, however, the deluded subject shows little or no tendency to bring the content *p* of their delusion into line with their belief set, and this puts pressure on the idea that they hold fast to *p* on the basis of sufficient subjective reasons for *p*. Next, we interrogate the notion of subjective reasons itself: how *subjective* can a subjective reason be? To have sufficient subjective epistemic reason for *p* is to be satisfied that one has uncovered sufficient objective reason for believing that *p*; now, the deluded subject certainly feels a kind of satisfaction, but should this be described as a satisfaction that there are sufficient objective reasons for the delusional content *p*? We will suggest that extreme cases of delusion put pressure upon this idea. Finally, we will draw attention to the fact that in extreme cases of delusion, we do not apply to the deluded subject the type of responsibility that we would need to in order to regard their state of delusion as subject to the correctness condition, and given that the correctness condition is constitutive of belief, it follows that we do not regard the deluded subject in such extreme cases as holding a belief. All-in-all, we will treat these considerations as establishing a strong case for the claim that extreme cases of delusion do not involve belief, and we will close with a brief review of some of the options available to account for such cases in non-doxastic terms.

6.1 Introducing Delusion

Roughly what kind of phenomenon is delusion? Unlike self-deception, the ascription of which is often made to those who would be understood as within a normal psychological range, delusions are considered symptomatic of psychiatric disorders, such as schizophrenia, and dementia, and characteristic of delusional disorders (Bortolotti, 2009, p. 1). The current *Diagnostic and Statistical Manual of Mental*

---

[33] As will become clear, the sense of *integration* at play here is unrelated to the sense of *integration* involved in chapter three, when we addressed Owens' (2000) denial that the believer aims at truth.

Disorders – the DSM – IV – published by the *American Psychiatric Association* defines delusion as follows:

"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 credibility." (2000, p. 765)

Now, this definition has been challenged along multiple lines, neatly summarised by Max Coltheart (2007, p. 1043):

"Couldn't a true belief be a delusion, as long as the believer had no good reason for holding the belief? 2. *Do delusions really have to be beliefs* – might they not instead be imaginings that are mistaken for beliefs by the imaginer? 3. Must all delusions be based on inference? 4. Aren't there delusions that are not about external reality? 'I have no bodily organs' or 'my thoughts are not mine but are inserted into my mind by others' are beliefs expressed by some people with schizophrenia, yet are not about external reality; aren't these nevertheless still delusional beliefs? 5. Couldn't a belief held by all members of one's community still be delusional?" (italics added)

A further widespread criticism of the DSM definition is that it does not foreground the well-being of the deluded person – the fact that delusions are preoccupying and distressing for the deluded person and often severely disrupt their day-to-day functioning. In this chapter we will only be interested in the various ways in which the DSM definition can be challenged in so far as they impact upon 2.: "Do delusions really have to be beliefs?" We use the definition merely as a jumping-off point for our investigation.

Now, delusions are typically divided into *monothematic* and *polythematic* kinds. There are two standard examples of monothematic delusions that have been picked up by philosophers of mind in recent times: the *Capgras* delusion and the *Cotard* delusion. People suffering from Capgras typically claim that a loved one has been

replaced by an impostor of some kind. Here is the description of the experience of one Capgras patient:

"I knew her name and everything about her, yet she appeared strange, unreal, like a statue. I saw her eyes, her nose, her lips moving, heard her voice and understood what she said perfectly, yet I was in the presence of a stranger." (Marguerite Sechehaye, 1970, p. 36)

Those suffering from Cotard delusion, on the other hand, characteristically claim to be dead. Jules Cotard (1882) calls the delusion *le délire de négation*:

"She repeatedly stated that she was dead and was adamant that she had died two weeks prior to the assessment (i.e. around the time of her admission on 19/11/2004). She was extremely distressed and tearful as she related these beliefs, and was very anxious to learn whether or not the hospital she was in, was 'heaven'. When asked how she thought she has died, LU replied 'I don't know how. Now I know that I had a flu and came here on 19th November. Maybe I died of the flu.' Interestingly, LU also reported that she felt 'a bit strange towards my boyfriend. I cannot kiss him, it feels strange – although I know that he loves me'." (Ryan McKay and Lisa Cipolotti, 2007, p. 353)

Both the Capgras and Cotard delusions are monothematic: they involve just one theme.[34] By way of contrast, polythematic delusions involve more than one theme, where the themes can be interrelated: a classic example is persecutory delusions, where the person in question is deluded that everyone hates them and is out to get them.

A second important distinction that is usually drawn with respect to delusions is the distinction between *circumscribed* and *elaborated* delusions. Monothematic delusions are typically circumscribed, whereas polythematic delusions tend to be elaborated. A delusion is more or less circumscribed depending upon the extent to which it leads to the formation of further states whose content is related to the content of the initial delusion, and the pervasiveness of the effects on the deluded person's behaviour. For example, it is fairly common in Capgras that although the

---

[34] However, as Bayne and Pacherie (2004, p. 8) point out, to call Capgras and Cotard monothematic is a simplification: "monothematic delusions do not have quite the monothematicity that they are often presented as having."

sufferer is deluded that their loved one has been replaced by an impostor, they show little interest in attempting to find their real loved one. However, Capgras can become elaborated, such as when the deluded person develops paranoid thoughts regarding the content of their delusion – for example that the impostor is evil and intends to harm them (Bortolotti, 2009, pp. 8-9).

6.2 Typical Contents of Delusions

We will now begin to consider various typical features of delusions that pertain to whether we should attribute the deluded subject a belief in the contents of their delusion. The first of these is the kind of contents delusions tend to have. Consider delusions of *inter-metamorphosis* – that one person has changed into another – for example. One patient claimed that his mother changed into another person every time she put her glasses on (Karel De Pauw and Krystyna Szulecka, 1988). Now, the content of such a claim might strike one as incoherent – perhaps meaningless. Is it not nonsensical to claim that one's mother *herself* changes into another person under certain conditions? Conceiving of such contents as meaningless would then lead one to deny that the deluded individual believes them: one cannot have a belief with meaningless content, as such content is not a candidate for truth evaluation. Indeed, John Campbell (2001, p. 95) claims that:

"The key question is whether the deluded subject can really be said to be holding on to the ordinary meanings of the terms used."

The Cotard delusion provides us with another good example: it may well be thought to involve a performative contradiction as a dead person could not assert *I am dead*. And, again, this claim could be understood as not simply necessarily false, but evidencing a lack of a firm grasp on the concept of death. However, as Matthew Ratcliffe (2008. p. 168) says:

"We should be wary of enforcing overly strict semantic boundaries and jumping to the conclusion that utterances to the effect that one is dead are meaningless."

Now, this is not to imply that the deluded individual uses the word *dead* in some non-literal, metaphorical sense, but rather, as Bayne and Pacherie (2005, p. 182) say:

"The semantics of a word may reflect the fact that it is part of several scientific or folk-psychological domains. For 'death' these domains include biology, the subjective realm of consciousness, religious beliefs concerning life after death, supernatural entities etc. In his use of words, a delusional patient may break the rules governing one of these intersecting language games while retaining the meaning his words have in other language games."

Consider LU's case, described above, in which she asks if she is in heaven. Presumably she uses the word *dead* to refer to her bodily death, which she conceives of her consciousness as having survived, and hence she wonders at the realm this consciousness now inhabits. Now, this does not mean that one need have a fully worked-out theory to support one's conception of the content of one's belief, of course. The person suffering from delusions of inter-metamorphosis, for example, need not have a theory of personal identity, according to which the kind of transitions he describes are possible; he must simply have some less than fully determinate meaning in mind – such as that under certain circumstances (glasses off) his mother is present and under other circumstances (glasses on) someone other than his mother is present. So, even in cases of the most bizarre assertions, there is often some meaning attributable to the deluded subject, and hence the possibility of their believing the content about which they are deluded is not precluded just by the nature of such content. Rather than looking to the *content* of the delusion, therefore, we will need to look to the nature of the deluded subject's *relation* to that content.

6.3 Integration

The next consideration we will look to is the allegation that because the contents of delusions are often poorly integrated with the rest of the deluded subject's beliefs – especially in the case of monothematic, circumscribed delusions – they cannot be said to believe them.

Why does it matter whether a subject's beliefs exhibit some form of rational integration? The idea that they must is famously associated with Quine (e.g. 1960), and Davidson (e.g. 1973), who thinks of this coherence constraint as the most fundamental notion of rationality. Now, we will not explore the positions of these two thinkers – rather, we will simply note that we can accommodate the idea that some degree of rational coherence is important for belief attribution with the

explication of the idea that beliefs aim at truth that we have arrived at so far. Whilst it is of course the case that there can be a lot of conflict within one's total set of beliefs, some basic coherence is nonetheless required by the idea that beliefs aim at truth in the sense we have articulated. Minimally, for example, given that beliefs aim at truth, one cannot consider whether to believe that *p* and reach the conclusion that *p and ~p* – providing one accepts that the contents *p and ~p* cannot be true.

So, in certain cases of delusion where the lack of integration between the contents of the delusion and the rest of the deluded subject's beliefs is fairly subtle, the existence of such conflict need not preclude attributing them a belief in the contents of their delusion. We might even allow that people suffering from *anosognosia* (denial of illness) who frequently deny that their limbs are paralysed, and yet recognise on other occasions that they cannot carry the shopping (Berti *et al*., 1998, cited in Bortolotti, 2010, p. 65), just hold an extremely irrational belief that it is not the case that their limbs are paralysed. However, deluded patients sometimes exhibit a far more blatant lack of integration between their beliefs:

"In the course of the same interview, she clearly stated that her husband had died suddenly four years earlier and had been cremated (this was correct), but also that her husband was a patient on the ward in the same hospital." (Breen *et al*., 2000, p. 91, cited in Bortolotti, 2010, p. 64)

Furthermore, they often fail to see the need to *resolve* such explicit bad integration. For example, LU:

"…was asked about how she knew whether someone was dead. She answered that dead people are motionless, and don't speak. Then she was asked to explain why she could move and speak, and she acknowledged the tension between her thoughts about dead people and her thought that she was dead. Yet she did not give up the delusion that she was dead." (Bortolotti, 2010, p. 64)

Now, in such explicit cases of bad integration – particularly when no need to resolve the conflict is felt by the patient – it seems that they *cannot* be said to hold the relevant belief(s), as this would contradict the idea that it is constitutive of belief that it aims at truth in the manner we have defended. As Currie and Ian Ravenscroft (2002, p. 176) put it:

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

Now, at this point, the doxasticist might object that cognitive behavioural therapy is often used with some success to treat deluded patients, and such therapy is characterised by questioning the likelihood that the content of the patient's delusion is true given their other beliefs. This could be understood to suggest that the deluded subject does experience some "obligation to resolve… inconsistencies between [their] beliefs" (Currie and Ravenscroft, 2002, p. 176)

However, even given that cognitive behavioural therapists *treat* deluded patients as if they believe the contents of their delusion, and have some success in so doing, the question is whether this implies that the deluded subject *does* believe the contents of their delusion. It is clear that it does not entail this: a form a therapy with an incorrect presupposition may be somewhat successful nevertheless. Indeed, it may be beneficial for a deluded subject to be treated as if they believe their delusion; it might even be that there will be a point during the therapy at which their delusional relation to the content *p* becomes an irrational belief that *p* even if it were not beforehand. None of this establishes anything about the status of the deluded subject's relation to the delusional content before they have stepped into the therapist's office. In any case, the success of cognitive behavioural therapy with deluded patients is limited, as Bortolotti (2010, p. 68) describes:

"Literature on the use of cognitive behavioural therapy for subjects with persecutory delusions suggests that the conviction in the content of the delusion and the preoccupation with the delusion can be reduced significantly as a result of these challenges, but the outcome does not often include a reassessment of the delusional states or a satisfactory elimination of dissonance."

Given these considerations, appeals to the manner in which cognitive behavioural therapists treat deluded patients do not offer the doxasticist's case much support: in extreme cases, it seems that the deluded subject does not exhibit even the minimal tendency towards integration that our account requires of a believer.

6.4 Subjective Reasons

A further consideration against doxasticism emerges if we think in more detail about what it *is* to possess a subjective reason. As we have said above, the possession of a subjective reason involves being in a state of satisfaction; but this state of satisfaction is not, as it were, untethered: one is satisfied that something in particular is the case, namely that one has uncovered sufficient objective reason for *p* in order for the belief that *p* to be rational. Now, the subjective reasons that one possesses – that is, the objective reasons one takes oneself to have discovered – and the objective reasons in fact available to one may depart very far from one another. Even in such cases of irrational belief, however, it is one's satisfaction that one has uncovered sufficient objective reasons – however misplaced – that marks one out as believer, as a rational agent, however inept one's execution of such agency may have turned out to be on the present occasion. Nonetheless, it seems that there are limits on how far one can go wrong while still taking oneself to have uncovered objective reasons. When one crosses this boundary, one will no longer qualify as being satisfied that one has uncovered objective reasons, and this is to say that one will no longer count as possessing subjective reasons, even if it is possible to undergo a sense of satisfaction that leaves one feeling, from the inside, just like somebody who does in fact possess subjective reasons.

This boundary has been crossed when it is so obvious that two things do not stand in a reasons relation to one another, that a subject who claims to be satisfied that they in fact do stand in such a relation is most naturally understood as failing to understand what it *is* for two things to stand in a reasons relation. Now, in certain extreme cases of delusion, it seems that the attempt to cast the deluded subject as believing the content about which they are deluded pushes them over exactly this boundary. Consider the following case of *erotomania* (Jordan *et al.*, 2006, p. 787, cited in Bortolotti 2010, p. 122):

"She realized he was empty without her and was pursuing her, but enemies were preventing them from uniting. The enemies included a number of people: people in her family, her classmates, neighbours and many other persons who were plotting to keep them apart. She knew that her conclusions were accurate because he would send messages to her proving his love. These messages would often present

themselves as the license plates on cars of a certain state, the color purple and other indications that she received from the environment that proved that he loved her."

In this example, the woman involved treats number plates and the presence of certain colours as *indicating* that a person loves her. But is this experience of indication sufficient for us to say of the woman that she is satisfied, on the basis of the fact that she sees certain number plates and colours, that she has available to her sufficient objective reasons to rationally form the belief that the person in question loves her? Surely not. It is so obvious that the presence of certain colours and the feelings of one person towards another ordinarily have no bearing upon one another that, were the woman to choose to express herself by saying *There's purple here, therefore he loves me*, we would want to question whether she understands, on this occasion, what it is for one thing to be a reason for another. Of course, especially in the case of circumscribed, monothematic delusions, the woman may have a large independent body of beliefs that remain untouched by her delusional state, and it is clear that her holding such beliefs testifies to a firm grasp of the nature of a reasons relation in other contexts. This does nothing to undermine our verdict on the present case, however. We are suggesting that the blatant outlandishness of the connections that the woman draws in her delusional state invites a verdict according to which her delusion involves a periodic alienation from her rational agency, rather than just a periodic lapse into irrationality.

In any case, the woman herself does not explicitly advert to the language of rational grounding when she explains the connections that she experiences between the colour purple and the person's love for her. She does not say, for instance, that the presence of the colour purple probabilifies, or suggests that the person loves her; it is rather the *indication* of his love – a *sign* that the environment has chosen to disclose only to her. There is thus a sense in which the connections that the woman encounters are *magical*, rather than rational. Her experience as of looking on at a world in which one thing can *indicate* another without exhibiting any of the intelligible connection characteristic of a reasons relation, may produce a feeling in her that superficially resembles the sense of consciously coming to be satisfied that one has sufficient objective reasons for rational belief that p (she claims that her convictions are *accurate* and *proven*). Regardless of any such phenomenological

affinity, however, we should not regard the woman's delusion as upheld on the basis of sufficient subjective reasons.

Now, this is not to assert that the deluded woman *must* have left the domain of the rational in order to treat the colour purple as testifying to the fact that the person in question loves her. If, for example, she tells herself that her loved one, or some representative of him, is in the local vicinity, and they are engineering the fact that she so frequently encounters the colour purple (knowing, perhaps, that it is her favourite colour), then, against this backdrop, we would be able to regard her as grasping the concept of a reasons relation if she asserted *There's purple here, therefore he loves me*. If she thinks in this way, then her delusion will have left the realm of the magical and re-entered the rational: her delusion will now qualify as an irrational belief. Nonetheless, absent such rationalisation, her delusion will not be the product of her having sufficient subjective reasons, because absent such rationalisation, it will not make sense to think of her as possessing subjective reasons.

6.5 The Deluded Subject and the Correctness Condition

This brings us to our final consideration regarding whether the deluded subject believes the content of their delusion. This consideration stems from the very fact that it is typical to refer to someone in such circumstances as a deluded *subject*, and that they are often – quite literally – treated as a *patient*. What this suggests is of course that the person is not to be held responsible for holding fast to the contents of their delusion. And, as we have argued, it is characteristic of belief that one is held rationally responsible for it. So, it seems that we may have hit upon a neat way of drawing a boundary between those with delusions that believe their delusions, and those who do not. That is, when the consensus is that the deluded subject is not to be held responsible for holding fast to the contents of their delusion, they are not to be understood as believing its contents. When, on the other hand, they continue to be seriously engaged with at the level of reasons, they ought to be attributed the belief in the contents of their delusion.

6.6. Non-Doxastic Accounts of Delusion

Given that we have argued that in certain cases deluded subjects fall outside of the bounds of belief attribution with respect to the contents of their delusion, we must establish that there are viable alternatives to the conception of delusion as involving belief. We will not be concerned to adjudicate between or argue for any particular such position. Nonetheless, we will now point to two *prima facie* viable options.

First, Velleman (2000, p. 266, fn 43) claims, that the deluded subject has a *phantasy* that *p*, as opposed to believing that *p*, where phantasies are understood as:

"…imaginings that are generated by wish rather than purposely formed by the subject out of a desire to imagine."

As he puts it:

"Aren't there people who believe that they are Napoleon? (People other than Napoleon, I mean.)…

I think the answer is that it isn't literally a belief. I suspect that we tend to apply the term 'belief' in a figurative sense to phantasies for which the subject doesn't or cannot have countervailing beliefs. When someone is said to believe that he is Napoleon, he actually has a phantasy to that effect; but on the question of who he is, a phantasy is all he has. He is somehow incapable of reality-tested cognitions of his identity. The phantasy of being Napoleon is thus what he has instead of a belief about his identity; and in this sense it is his belief on the topic, just as a cardboard box on the sidewalk may be his house in virtue of being what he has instead of a house.

If you ask me, however, a cardboard box on the sidewalk isn't really a house. And a phantasy of being Napoleon isn't really a belief." (Velleman, 2000, pp. 280-281)

So, Velleman thinks of the subject who is deluded that *p* as imagining that *p*, as opposed to believing that *p*. Such an analysis seems plausible in the light of the fact that deluded subjects – particularly in circumscribed cases – often fail to act consistently in accordance with the contents of their delusion. Think of the Capgras patient who claims that their loved one has been replaced by an impostor, but shows little concern regarding the whereabouts of their *real* loved one. This would be well-

explained by the fact that they imagine, rather than belief their loved one to be missing – whilst my belief that my husband is missing would ordinarily dispose me to go out and look for him, my imagining that this is so would not.[35] Velleman's account would also neatly explain the fact that, in extreme cases, the deluded subject exhibits little or no tendency to integrate the deluded content with their other beliefs, since one is certainly not obligated to integrate one's imaginings with one's beliefs.

A second, related, account that could also lay claim to both of these explanatory benefits is that developed by Currie (2000), known as the *metarepresentational* account. On Currie's account, like Velleman's, the deluded subject has nothing but an imagining where others would have a belief; but, moreover, according to Currie the subject *misidentifies* their imagining that *p* as a belief that *p*. Thus, Currie is well placed to explain the deluded subject's behaviour and lack of integration in terms of the dispositional profile of an imagining, and indeed his account is explicitly designed to show how:

"…it would be intelligible how it is that people with delusions often do not follow out the consequences of their delusions very far, or try to resolve tensions between their delusions and their other beliefs." (Currie and Ravenscroft, 2002, p. 178)

However, in addition to explaining these features, the metarepresentational account could also be capable of explaining the additional items of behaviour – the deluded subject's conviction that *p*, for example – that superficially point towards their holding a belief.

Conclusion

In this final chapter, we have considered what our conception of belief has to teach us about delusion and *vice versa*. We began by introducing the phenomenon of delusion, with particular reference to the Capgras and Cotard delusions. We then proceeded to explore four main considerations pertaining to whether or not the deluded subject believes the content of their delusion. The first of these was the contention that the deluded subject cannot be understood to believe the contents of their delusion, because these contents are meaningless, or because their assertion

---

[35] It should be noted that Velleman himself would not approve this use of his account, claiming as he does that imagination and belief have an indistinguishable motivational role. See O'Brien (2005a) for a convincing critique of this claim.

amounts to a performative contradiction – as in the case of the Cotard patient who asserts *I am dead*. However, we argued that we ought not to be so literal-minded in interpreting the meaning of such patients. Given that the adjective *dead* can be understood in different ways in accordance with different linguistic and cultural practices, the patient's assertion clearly need not be meaningless. For example, they might think of their consciousness as having survived their bodily death. So, we maintained that such accusations of meaninglessness do not preclude us from thinking of the deluded subject as believing the contents of their delusion. Next, we turned to the claim that the deluded person cannot be said to believe the contents of their delusion because such a belief would be too poorly integrated with the rest of their beliefs. We acknowledged that, although a certain amount of undiscovered inconsistency in someone's belief system is to be accepted, if a deluded person asserts propositions that are evidently contradictory in the same breath, and then sees no need to resolve such a contradiction, they have exceeded our capacity to attribute them beliefs. Given that beliefs aim at truth as we have argued, one cannot be understood to believe that *p and ~p*. We then addressed in more detail what it is to possess a subjective reason, and argued that there are certain extreme cases of delusion in which we could just as soon regard the deluded subject as failing to grasp or apply the concept of a reasons relation as see them as irrationally operating on the basis of extremely idiosyncratic subjective reasons. In such cases, we argued, delusion constitutes a periodic alienation from one's rational agency, rather than simply a brief foray into the domain of the irrational, and for this reason we suggested that the concept of belief does not get a grip on such cases. Finally, we argued that the fact that the attribution of delusion often carries with it the attribution of patiency is itself a barrier to attributing the deluded subject a belief in the contents of their delusion. As we have argued, belief is a fundamentally normative concept, for which one is to be held directly rationally responsible.

Overall Conclusion

In this thesis, we set out to unearth some significant features of our folk psychological concept of belief, and this we have done. We have discovered that beliefs aim at truth in the sense that:

(R): It is constitutive of belief that *if* it is consciously regulated, it is so-regulated solely for truth;

and

(C): It is constitutive of belief that it is correct if and only if its contents are true.

We have argued that (C) represents the ultimate explanation of (R), and hence that belief is at base a normative phenomenon, for which the question as to why it is held can always be raised. In the process, we have highlighted the fact that folk psychology is rich in potentially explanatory concepts, and that if we allow concepts other than belief and desire to do their own work, we remove the temptation to misunderstand belief and its relationship with reasons.

We began by considering two cases in which it is difficult to attribute belief: Implicit Racist and Precipice. We argued that in order to explain the patterns in the behaviour of both Juliet and the woman in Precipice, it seems that we need to postulate that there is a significant relationship between belief and reasons. We then began to investigate this relationship.

In chapter two we investigated two features of belief that appear to point in opposing directions. The first is the transparency of doxastic deliberation – that when one considers *whether to believe that p* this is immediately transparent to the question *whether p*. This seems to suggest that belief and reasons are importantly intertwined. On the other hand, however, heterogeneity – the idea that beliefs are often influenced (outside of conscious awareness) by factors other than epistemic reasons – seems to suggest that the relationship between belief and reasons might not be all that significant after all. However, we argued that we can explain transparency – in spite of heterogeneity – by appealing to the fact that in considering *whether to believe that p*, the doxastic deliberator understands that the only considerations relevant to this question are those that determine *whether p*, because they understand that it is

constitutive of belief that it is answerable solely to epistemic norms – ultimately the correctness condition. They understand that it is because of (C) that (R) is the case: in consciously regulating their state, they would not be consciously regulating a belief unless they were doing so solely in accordance with truth.

Having thus explained transparency and argued for both (R) and (C), we then set out to explore each of these conditions in further detail. We began in chapter three with the regulation condition. We argued that it is not to be understood in terms of the doxastic deliberator themselves possessing an aim or intention to believe. We cannot maintain this to be the case because in attempting to do so, we impale ourselves upon either one or another horn of a dilemma. On one horn, we attempt to model the doxastic deliberator's aim as to believe that p if and only if p. If we do this, however, their deliberation as to *whether to believe that p* would fail to exhibit the *exclusivity* it in fact does. That is, because one can be uncertain as to *whether p*, one's aim to believe that p if and only if p could *integrate* with one's other aims in the following sense. One could think: *Well, I've got quite good evidence that p and, given that time is short, I'll go ahead and form the belief that p*. This cannot happen. So, we might try to model the doxastic deliberator's aim as follows: to believe that p if and only if they are *satisfied* that p. However, this will impale us upon the other horn of the dilemma, as such an aim would be *circular*: to be consciously satisfied that p simply *is* to consciously believe that p. By way of conclusion, we maintained that, according to the regulation condition, beliefs aim at truth not in the sense that the believer aims at truth, but rather in virtue of the fact that the belief *itself* aims at truth. The belief itself aims at truth in the sense that the believer's coming to hold it *is* their coming to be satisfied that it is true.

In chapter four, we turned to the correctness condition: that the belief that p is correct if and only if p. We argued that from this correctness condition, we can derive two further epistemic norms – normative exclusivity and demandingness – which combine to yield:

(Ra): One should believe that p if and only if one does so solely on the basis of sufficient objective epistemic reason in favour of p.

We considered two main objections to this idea. The first is the *pragmatist*'s claim that belief is answerable to practical norms. We argued that the pragmatist fails to

distinguish two questions: *whether to believe that p* and *whether to try to get oneself to believe that p*; whilst the latter is answerable to practical norms, the former is answerable solely to epistemic norms. We then considered a second objection to the claim that belief is a fundamentally normative concept, which has it that we cannot be understood to have the requisite kind of control over our beliefs in order to be held responsible for them in the manner in which we are. We engaged with Owens' (2000) forceful articulation of this objection, and argued that he has not established that we do not possess a kind of *reflective control* over our beliefs that would make sense of our practices of holding one another responsible in a manner consistent with (C).

In chapters five and six, we came full circle and considered two phenomena – self-deception and delusion – which present difficulties for belief attribution, just as Implicit Racist and Precipice did at the start. Our hope was that, in so doing, we would learn more both about the concept of belief we have been developing, as well as about self-deception and delusion themselves. We began in chapter five with self-deception. We argued for a position we called *doxastic minimalism*. According to the doxastic minimalist, in the paradigm case of self-deception, the self-deceiver neither believes that *p* – their undesired belief – nor believes that ~*p* – their desired belief. We began by presenting a negative argument that this is the case. We argued that, considered individually, no explanatory need to attribute the self-deceiver either of their purported beliefs has been demonstrated. Facts about their behaviour and the phenomenology of self-deception can equally well be explained by appeal to other folk psychological concepts, such as *suspicion*, *anxiety*, and *hope*. We then presented a positive case in favour of our *doxastic minimalist* conclusion. In the case of the self-deceiver's undesired belief, we argued that they cannot be said to hold such a belief, because if they did, they would have to do so unconsciously in the strong sense that they would be unable to reflect upon it. This conflicts with our claim that we can be held directly responsible for our beliefs in the manner in which we are insofar as we can reflect upon them at given time. So, in considering self-deception we revealed that a consequence of our account of belief is that there cannot be unconscious beliefs, in the strong sense. Regarding the self-deceiver's desired belief, we argued that they cannot be understood to hold such a belief, as when they consciously think about ~*p* they do not do so in a manner that can be understood to

constitute conscious regulation solely in accordance with truth, as they experience a characteristic unease surrounding the topic *p* that is incompatible with such regulation.

In chapter six, we turned to the phenomenon of delusion. As with self-deception, our interest in delusion was an interest in belief attribution: specifically, we were concerned to determine whether or not the deluded subject can be said to *believe* the content about which they are deluded. Whilst we acknowledged that a number of central cases are plausibly regarded as species of irrational beliefs, we denied that such a *doxasticist* construal of delusion adequately accounts for certain more extreme cases. In such cases, the subject is better seen as periodically alienated from their nature as a rational agent and their status as a believer, rather than forming an irrational belief. Three considerations motivated this verdict: first, subjects in such cases show little or no tendency to rationally integrate the delusional content with their other beliefs, secondly, it puts a strain on our concept of a reason to describe them as taking themselves to have sufficient reason in support of the delusional content, and finally, we do not extend to such subjects the kind of rational agency that is required for us to hold them rationally responsible, and hence for the correctness condition to hold. Thus, delusion afforded a clarification and affirmation of the conception of belief developed in the first four chapters. Through attention to the deluded subject, we re-learned that belief is a fundamentally normative notion, and that as soon as one thinks of oneself as forming a belief, one thinks of oneself as responding to its constitutive norms. Thus, to believe is to claim for oneself a kind of rational agency that neither the self-deceiver nor the deluded subject is fully capable of living out.

Bibliography

Adler, Jonathan 2002: *Belief's Own Ethics*, Cambridge, MA: The MIT Press.

American Psychiatric Association 2000: *Diagnostic Statistical Manual of Mental Disorders* (4th edn), Text Revision (DSM-IV-TR).

Aristotle 1998: *The Nicomachean Ethics*, Ross, D. (trans.), Oxford: Oxford University Press.

Armstrong, David 1968: *A Materialist Theory of the Mind*, London: Routledge & K. Paul.

Audi, Robert 1972: 'The Concept of Believing', *The Personalist*, 53, pp. 43-62.

Audi, Robert 1982: 'Self-Deception, Action, and Will', *Erkenntnis*, 18 (2), pp. 133-158.

Audi, Robert 1988: 'Self-Deception, Rationalization, and Reasons for Acting', *Perspectives on Self-Deception*, McLaughlin, B. and Rorty, A. (eds), Berkeley: University of California Press, pp. 92-120.

Audi, Robert 1997: 'Self-Deception vs. Self-Caused Deception: A Comment on Professor Mele', *Behavioural and Brain Sciences*, 20, p. 104.

Audi, Robert 2007: 'Belief, Intention, and Reasons for Action', *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi*, Timmons, M., Greco, J., and Mele, A. (eds), Oxford: Oxford University Press, pp. 248-262.

Austin, John 1961: *Philosophical Papers*, Urmson, J. and Warnock, G. (eds), New York: Oxford University Press.

Baker, Lynne 1999: 'Folk Psychology', *MIT Encyclopedia of Cognitive Science*, Wilson R. and Keil, F. (eds), Cambridge, MA: MIT Press, pp. 317-318.

Barnes, Annette 1997: *Seeing Through Self-Deception*, Cambridge: Cambridge University Press.

Bayne, Tim and Pacherie, Elisabeth 2004: 'Bottom-up or Top-down? Campbell's Rationalist Account of Monothematic Delusions', *Philosophy, Psychiatry & Psychology*, 11, pp. 1-11.

Bayne, Tim and Pacherie, Elisabeth 2005: 'In Defence of the Doxastic Conception of Delusions', *Mind & Language*, 20 (2), pp. 163-188.

Bayne, Tim 2010: 'Delusions as Doxastic States: Contexts, Compartments, and Commitments', *Philosophy, Psychiatry, and Psychology*, 17 (4), pp. 329-336.

Bermúdez, José Luis 2000: 'Self-Deception, Intentions and Contradictory Beliefs', *Analysis*, 60 (4), pp. 309-319.

Bermúdez, José 2001: 'Normativity and Rationality in Delusional Psychiatric Disorders', *Mind & Language*, 16 (5), pp. 457–493.

Boghossian, Paul 2003: 'The Normativity of Content', *Philosophical Issues*, 13, pp. 31-45.

Bortolotti, Lisa 2009: 'Delusion', *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/*entries/delusion/*

Bortolotti, Lisa 2010: *Delusions and Other Irrational Beliefs*, Oxford: Oxford University Press.

Boyle, Matthew 2009: 'Two Kinds of Self-Knowledge', *Philosophy and Phenomenological Research*, 78 (1), pp. 133-164.

Boyle, Matthew 2011a: ''Making Up Your Mind' and the Activity of Reason', *Philosophers' Imprint*, 11 (17), www.philosophersimprint.org/011017/

Boyle, Matthew forthcoming a: 'Active Belief', *Canadian Journal of Philosophy*, 41, http://nrs.harvard.edu/urn-3:HUL.InstRepos:4875850

Boyle, Matthew forthcoming b: 'Essentially Rational Animals', *Rethinking Epistemology*, Abel, G. and Conant, J. (eds), Berlin: Walter de Grutyer, http://nrs.harvard.edu/urn-3:HUL.InstRepos:8641838

Bratman, Michael 1992: 'Practical Reasoning and Acceptance in a Context', *Mind*, 101 (401), pp. 1-15.

Byrne, Alex and Gertler, Brie 2010: 'Self-Knowledge of Beliefs', *Philosophy TV*, http://www.philostv.com/alex-byrne-and-brie-gertler/

Byrne, Alex and Boyle, Matthew 2011: 'Self-Knowledge and Transparency', *Proceedings of the Aristotelian Society Supplementary Volume*, 85 (1), pp. 201-241.

Campbell, John 2001: 'Rationality, Meaning, and the Analysis of Delusion', *Philosophy, Psychiatry, & Psychology*, 8 (2-3), pp. 89-100.

Cavell, Marcia 1994: 'Reason and the Gardener', *The Philosophy of Donald Davidson*, Hahn L. (ed.), Illinois: Open Court Publishing Company, pp. 407-421.

Chignell, Andrew 2010: 'The Ethics of Belief', *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/ethics-belief/

Chinn, Clark, and Brewer, William 2001: 'Models of Data: A Theory of How People Evaluate Data', *Cognition and Instruction*, 19 (3), pp. 323-393.

Chrisman, Matthew 2008: 'Ought to Believe', *The Journal of Philosophy*, 105 (7), pp. 346-370.

Chrisman, Matthew 2010: 'The Aim of Belief and the Goal of Truth: Reflections on Rosenberg', *Self, Language, and World: Problems from Kant, Sellars, and Rosenberg*, O'Shea, J. and Rubenstein, E. (eds), Atascadero, CA: Ridgeview Publishing Company.

Churchland, Paul 1981: 'Eliminative Materialism and the Propositional Attitudes', *Mind and Cognition*, Lycan, W. (ed.), Oxford: Blackwell, pp. 206-223.

Clifford, William 1999: 'The Ethics of Belief', *The Ethics of Belief and Other Essays*, Madigan, T. (ed.), Amherst, MAL Prometheus, pp. 70-96.

Cohen, L. Jonathon, 1992: *An Essay on Belief and Acceptance*, Oxford: Clarendon Press.

Cotard, Jules 1882: 'Du Délire des Négations', *Archives de Neurologie*, 4, pp. 152–170, 282–95.

Coltheart, Max 2007: 'Cognitive Neuropsychiatry and Delusional Belief', (The 33rd Sir Frederick Bartlett Lecture), *The Quarterly Journal of Experimental Psychology*, 60 (8), pp. 1041-1062.

Currie, Gregory 2000: 'Imagination, Delusion, and Hallucinations', *Pathologies of Belief*, Coltheart, M. and Davies, M. (eds), Oxford: Blackwell, pp. 167-182.

Currie, Gregory and Ravenscroft, Ian 2002: *Recreative Minds: Imagination in Philosophy and Psychology*, New York: Oxford University Press.

Davidson, Donald 1969: 'How is Weakness of the Will Possible?', *Essays on Actions and Events*, Oxford: Clarendon Press, pp. 21-42.

Davidson, Donald 1973: 'Radical Interpretation', *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, pp. 125-140.

Davidson, Donald 1982: 'Paradoxes of Irrationality', *The Essential Davidson*, Oxford: Oxford University Press, pp. 138-152.

Davidson, Donald 1984: *Inquiries into Truth and Interpretation*, Oxford: Clarendon.

Davidson, Donald 1986: 'Deception and Division', *Problems of Rationality*, Oxford: Clarendon Press.

Davidson, Donald 1997: 'Who is Fooled?', *Problems of Rationality*, Oxford: Clarendon Press.

Dennett, Daniel 1991: 'Real Patterns', *The Journal of Philosophy*, 88 (1), pp. 27-51.

Dennett, Daniel 1995b: 'Do Animals Have Beliefs?', *Comparative Approaches to Cognitive Science*, Roitblat, H. (ed.), Cambridge, MA: MIT Press.

De Pauw, Karel and Szulecka, Krystyna 1988: 'Dangerous Delusions: Violence and the Misidentification Syndromes' *British Journal of Psychiatry*, 152, pp. 91-96

De Rose, Keith 1992: 'Contextualism and Knowledge Attributions', *Philosophy and Phenomenological Research*, 52, pp. 913-929.

Descartes, René 1984: *The Philosophical Writings of Descartes Volume II,* Cottingham, J., Stoothoff R., and Murdoch, D. (trans.), Cambridge: Cambridge University Press.

Descartes, René 1985: *The Philosophical Writings of Descartes Volume I*, trans. Cottingham, J., Stoothoff R., and Murdoch, D. (trans.), Cambridge: Cambridge University Press.

Dupuy, Jean-Pierre (ed.) 1998: *Self Deception and Paradoxes of Rationality*, Stanford: CSLI Publications.

Egan, Andy 2008: 'Imagination, Delusion, and Self-Deception', *Delusion and Self-Deception: Affective and Motivational Influences on Belief*, Bayne, T. and Fernández, J. (eds), Hove: Psychology Press, pp. 263-280.

Elster, Jon 1999: 'Davidson on Weakness of Will and Self-Deception', *The Philosophy of Donald Davidson*, Hahn, L. (ed.), Illinois: Open Court Publishing Company, pp. 425-442.

Evans, Gareth 1982: *The Varieties of Reference*, McDowell, J. (ed.), Oxford: Clarendon Press.

Evans, Jonathan St. B. T. 2003: 'In Two Minds: Dual-Process Accounts of Reasoning', *TRENDS in Cognitive Science*, 7 (10), pp. 454-459.

Evans, Jonathan St. B. T. 2008: 'Dual-Processing Accounts of Reasoning, *Judgment and Social Cognition*, http://www.open.ac.uk/Arts/philos/mmr/Evans_ARP_dual_processes.pdf

Fotopoulou, Aikaterini, 2008: 'False-Selves in Neuropsychological Rehabilitation: The Challenge of Confabulation', *Neuropsychological Rehabilitation*, 18 (5/6), pp. 541-565.

Frankish, Keith 2004: *Mind and Supermind*, Cambridge: Cambridge University Press.

Frankish, Keith 2007: 'Deciding to Believe Again', *Mind*, 116, pp. 523-547.

Frankish, Keith 2012: 'Delusions, Levels of Belief, and Non-doxastic Acceptances', *Neuroethics*, 5, pp. 23-27.

Funkhouser, Eric 2005: 'Do the Self-Deceived Get What They Want?', *Pacific Philosophical Quarterly*, http://comp.uark.edu/~efunkho/selfdeception.pdf

Funkhouser, Eric 2009: 'Self-Deception and the Limits of Folk Psychology', *Social Theory and Practice*, http://comp.uark.edu/~efunkho/selfdecrat.pdf

Gardner, Sebastian 1993: *Irrationality and the Philosophy of Psychoanalysis*, Cambridge: Cambridge University Press.

Gendler, Tamar 2008a: 'Self-Deception as Pretense', *Philosophical Perspectives: Mind*, 21, pp. 231-258.

Gendler, Tamar Szabó 2008b: 'Alief and Belief', *The Journal of Philosophy*, pp. 634-663.

Gendler, Tamar Szabó 2008c: 'Alief in Action (and Reaction)', *Mind and Language*, 23 (5), pp. 552-585.

Gibbard, Allan 2005: 'Truth and Correct Belief', *Philosophical Issues*, 15, pp. 338-350.

Goldman, Alvin I. 1993: 'The Psychology of Folk Psychology', *Behavioural and Brain Sciences*, 16,
http://users.ecs.soton.ac.uk/harnad/Papers/Py104/goldman.psyc.html

Gorman, Michael 1993: 'Hume's Theory of Belief', *Hume Studies*, 19 (1), pp. 89-102.

Griffiths, Paul 2009: 'The Distinction Between Innate and Acquired Characteristics', *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/innate-acquired/

Halligan, Peter and Marshall, John 1996: 'The Wise Prophet Makes Sure of the Event First: Hallucinations, Amnesia, and Delusions', *Method in Madness*, Halligan P. and Marshall J. (eds), Exeter: Psychology Press, pp. 235-266.

Hieronymi, Pamela 2008: 'Responsibility for Believing', *Synthese*, 161, pp. 357-373.

Holton, Richard 2001: 'What is the Role of the Self in Self-Deception?', *Proceedings of the Aristotelian Society*, 101 (1), pp. 53-69.

Hume, David 1975: *An Enquiry Concerning Human Understanding*, Selby-Bigge, L. (ed.), Oxford: Oxford University Press.

Hume, David 1978: *A Treatise of Hume Nature*, Selby-Bigge, L. (ed.), Oxford: Oxford University Press.

Hunter, David 1998: 'Understanding and Belief', *Philosophy and Phenomenological Research*, 58 (3), pp. 559-580.

Hunter, David 2011: 'Alienated Belief', *dialectica*, 65 (2), pp. 221-240.

Jackson, Frank and Pettit, Philip 1993: 'Folk Belief and Commonplace Belief', *Mind and Language*, 8 (2), pp. 298-305.

James, William 1931: 'The Will to Believe', *The Will to Believe and Other Essays in Popular Philosophy*, London: Longmans, Green and Co.

James, William 1983: *The Principles of Psychology*, Cambridge, MA: Harvard University Press.

Johnston, Mark 1988: 'Self-Deception and the Nature of Mind', *Perspectives on Self-Deception*, McLaughlin, B. and Rorty A. (eds), Berkeley: University of California Press, pp. 63-91.

Kolodny, Niko 2007: 'How Does Coherence Matter?', *Proceedings of the Aristotelian Society*, 107 (3), pp. 229-263.

Korcz, Keith 2010: 'The Epistemic Basing Relation', *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/basing-epistemic/

Lazar, Ariela 1998: 'Akrasia and the Principle of Continence or What the Tortoise Would Say to Achilles', *The Philosophy of Donald Davidson*, Hahn, L. (ed.), Illinois: Open Court Publishing Company, pp. 381-401.

Lehrman, Daniel 1953: 'Critique of Konrad Lorenz's Theory of Instinctive Behavior', *Quarterly Review of Biology*, 28 (4), pp. 337-363.

Locke, John 1996: *An Essay Concerning Human Understanding*, Winkler, K. (ed.), Indiana: Hackett Publishing Company, Inc.

Lorenz, Konrad 1965: *Evolution & the Modification of Behaviour*, Chicago: University of Chicago Press.

Malcolm, Norman 1967: 'Knowledge and Belief', *Knowledge and Belief*, Griffiths, A. Phillips (ed.), Oxford: Oxford University Press, pp. 69-81.

Marler, Peter 2004: 'Science and Birdsong: The Good Old Days', *Nature's Music: The Science of Birdsong*, Marler, P. and Slabbekoorn, H. (eds), Elsevier Academic Press, San Diego , CA , pp. 1-38.

Matthews, Eric 2005: 'Unconscious Reasons', *Philosophy, Psychiatry, & Psychology*, 12 (1), pp. 55-57.

McDowell, John 1994: *Mind and World*, Cambridge, MA: Harvard University Press.

McDowell, John 2009: *Having the World in View: Essays on Kant, Hegel, and Sellars*, Cambridge, MA: Harvard University Press.

McHugh, Conor 2011a: 'Belief and Aims', *Philosophical Studies*, https://sites.google.com/site/conormchughhomepage/

McHugh, Conor 2011b: 'Epistemic Deontology and Voluntariness', *Erkenntnis*, https://sites.google.com/site/conormchughhomepage/

McHugh, Conor 2011c: 'What Do We Aim At When We Believe?', *dialectica*, 65 (3), pp. 369-392.

McHugh, Conor 2012a: 'The Truth Norm of Belief', *Pacific Philosophical Quarterly*, 93, pp. 8-30.

McHugh, Conor 2012b: 'The Illusion of Exclusivity', *European Journal of Philosophy*, https://sites.google.com/site/conormchughhomepage/

McKay, Ryan, Cipolotti, Lisa 2007: 'Attributional Styles in a case of Cotard Delusion', *Consciousness and Cognition*, 16, pp. 349-359.

McLaughlin, Brian 1988: 'Exploring the Possibility of Self-Deception in Belief', *Perspectives on Self-Deception*, McLaughlin, B. and Rorty, A. (eds), Berkeley: University of California Press, pp. 29-62.

Mele, Alfred 1982: ''Self-Deception, Action, and Will': Comments', *Erkenntnis*, 18 (2), pp. 159-164.

Mele, Alfred 1987a: *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*, Oxford: Oxford University Press.

Mele, Alfred 1992a: *Springs of Action: Understanding Intentional Behavior*, Oxford: Oxford University Press.

Mele, Alfred 1997: 'Real Self-Deception', *Behavioural and Brain Sciences*, 20, pp. 91-136.

Mele, Alfred 2001: *Self-Deception Unmasked*, Oxford: Princeton University Press.

Mele, Alfred 2007: 'Intention, Self-Deception, and Reasons for Action', *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi*, Timmons, M., Greco, J., and Mele, A. (eds), Oxford: Oxford University Press.

Mele, Alfred 2009: 'Have I Unmasked Self-Deception or Am I Self-Deceived?', *The Philosophy of Deception*, Clancy, M. (ed.), Oxford: Oxford University Press.

Mellor, Hugh 1978: 'Conscious Belief', *Proceedings of the Aristotelian Society*, 78, pp. 87-101

Mellor, Hugh 1980: 'Consciousness and Degrees of Belief', *Matters of Metaphysics*, Cambridge: Cambridge University Press, pp. 30-60.

Montaigne, Michel de 1957: *Essays,* Frame, D. (trans.), Palo Alto: Stanford University Press.

Moran, Richard 2001a: *Authority and Estrangement: An Essay on Self-Knowledge*, Oxford: Princeton University Press.

Moran, Richard 2001b: 'Responses to O'Brien and Shoemaker', *European Journal of Philosophy*, 11 (3), pp. 402-419.

Morton, Adam 1980: *Frames of Mind: Constraints on the Common-Sense Conception of the Mental*, Oxford: Clarendon Press.

Morton, Adam 1988: 'Partisanship', *Perspectives on Self-Deception*, McLaughlin, B. and Rorty A. (eds), Berkeley: University of California Press, pp. 170-182.

Morton, Adam 2003: *The Importance of Being Understood: Folk Psychology as Ethics*, London: Routledge.

Nichols, Shaun 2002: 'Folk Psychology', *Encyclopedia of Cognitive Science*, London: Nature Publishing Group, http://www.hum.utah.edu/philosophy/faculty/nichols/Papers/FolkPsychologyFinal.htm

Nickel, Philip 2010: 'Voluntary Belief on a Reasonable Basis', *Philosophy and Phenomenological Research*, 81, pp. 312-334.

O'Brien, Lucy 2001: 'Moran on Agency and Self-Knowledge', *European Journal of Philosophy*, 11 (3), pp. 375-390.

O'Brien, Lucy 2005a: 'Imagination and the Motivational View of Belief', *Analysis*, 65 (1), pp. 55-62.

O'Brien, Lucy 2005b: 'Self-Knowledge, Agency and Force', *Philosophy and Phenomenological Research*, 71 (3), pp. 580-601.

O'Brien, Lucy and Soteriou, Matthew, (eds), 2009: *Mental Actions*, Oxford: Oxford University Press.

Owens, David 2000: *Reason without Freedom: The Problem of Epistemic Normativity*, Routledge: Bury St Edmunds.

Owens, David 2003: 'Does Belief Have an Aim?', *Philosophical Studies*, 115 (3), pp. 283-305.

Pascal, Blaise 2006: *Pensées*, Project Gutenberg, http://www.gutenberg.org/catalog/world/readfile?fk_files=230692

Papineau, David forthcoming: 'There Are No Norms of Belief', http://kcl.academia.edu/DavidPapineau/Papers/968647/There_Are_No_Norms_of_Belief

Pears, David 1984: *Motivated Irrationality*, Oxford: Clarendon Press.

Price, Henry 1969: *Belief*, London: George Allen & Unwin.

Priest, Graham 2006: *In Contradiction: A Study of the Transconsistent*, Oxford: Clarendon Press.

Quine, Willard 1960: *Word and Object*, Cambridge, MA: The MIT Press.

Railton, Peter 1994: 'Truth, Reason, and the Regulation of Belief', *Philosophical Issues*, 5, pp. 71-93.

Ramsey, Frank 1926: 'Truth and Probability', *The Foundations of Mathematics and other Logical Essays*, Braithwaite, R. (ed.), London: Kegan, Paul, Trench, Trubner & Co., pp. 156-198.

Ratcliffe, Matthew 2008: *Feelings of Being: Phenomenology, Psychiatry and the Sense of Reality*, Oxford: Oxford University Press.

Reisner, Andrew and Steglich-Petersen, Asbjorn 2011: *Reasons for Belief*, Cambridge: Cambridge University Press.

Rosen, Gideon 2001: 'Brandom on Modality, Normativity and Intentionality', *Philosophy and Phenomenological Research*, 63 (3), pp. 611-623.

Rowbottom, Darrell Patrick 2007: '"In-Between Believing" and Degrees of Belief', *Teorema*, 26 (1), pp. 131-137.

Ryle, Gilbert 2000: *The Concept of Mind*, London: Penguin Books.

Schwitzgebel, Eric 2001a: 'On Containers and Content, with a Cautionary Note to Philosophers of Mind', http://www.faculty.ucr.edu/~eschwitz/SchwitzAbs/Containers.htm

Schwitzgebel, Eric 2001b: 'In-Between Believing', *The Philosophical Quarterly*, 51 (202), pp. 76-82.

Schwtizgebel, Eric 2002: 'A Phenomenal Dispositional Account of Belief', *Noûs*, 36 (2), pp. 249-275.

Schwitzgebel, Eric 2006: 'Belief', *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/belief/

Schwitzgebel, Eric 2008 'Mad Belief?', *The Splintered Mind*, http://schwitzsplinters.blogspot.com

Schwitzgebel, Eric 2010: 'Acting Contrary to Our Professed Beliefs, or the Gulf Between Dispositional Belief and Occurrent Judgment', *Pacific Philosophical Quarterly*, 91, pp. 531-553.

Sechehaye, Marguerite 1970: *Autobiography of a Schizophrenic Girl*, New York: Signet.

Shah, Nishi 2002: 'Clearing Space for Doxastic Voluntarism', *The Monist*, 85 (3), pp. 436-445.

Shah, Nishi 2003: 'How Truth Governs Belief', *The Philosophical Review*, 112 (4), pp. 447-482.

Shah, Nishi and Velleman, David 2005: 'Doxastic Deliberation', *The Philosophical Review*, 114 (4), pp. 497-534.

Shah, Nishi 2006: 'A New Argument for Evidentialism', *The Philosophical Quarterly*, 56 (225), pp. 481-498.

Shoemaker, Sydney 2001: 'Moran on Self-Knowledge', *European Journal of Philosophy*, 11 (3), pp. 391-401.

Smith, Michael 1994: *The Moral Problem*, Oxford: Blackwell Publishing.

de Sousa, Ronald 1988: 'Emotion and Self-Deception', *Perspectives on Self-Deception*, McLaughlin, B. and Rorty, A. (eds), Berkeley: University of California Press, pp. 324-341.

Steglich-Petersen, Asbjørn 2006a: 'Voluntarism and Transparent Deliberation', *South African Journal of Philosophy*, 25 (2), pp. 171-176.

Steglich-Petersen, Asbjørn 2006b: 'No Norm Needed: on the Aim of Belief', *The Philosophical Quarterly*, 56 (225), pp. 499-516.

Steglich-Petersen, Asbjørn 2008a: 'Against Essential Normativity of the Mental', *Philosophical Studies*, pp. 263-283.

Steglich-Petersen, Asbjørn 2008b: 'Does Doxastic Transparency Support Evidentialism?', *dialectica*, 62 (4), pp. 541-547.

Steglich-Petersen, Asbjørn 2009: 'Weighing the Aim of Belief', *Philosophical Studies*, 145, pp. 395-405.

Steglich-Petersen, Asbjørn 2011: 'How to be a Teleologist about Epistemic Reasons', *Reasons for Belief*, Reisner, A. and Steglich-Petersen, A. (eds.), Cambridge: Cambridge University Press.

Stocker, Michael 1982: 'Responsibility Especially for Beliefs', *Mind*, pp. 398-417.

Strawson, Galen 2003: 'Mental Ballistics or the Involuntariness of Spontaneity', *Proceedings of the Aristotelian Society*, 103, pp. 227-256.

Sullivan-Bissett, Ema and Noordhof, Paul 2011: 'A Defence of Owens' Exclusivity Objection to Beliefs having Aims', *Philosophical Studies*, http://rd.springer.com/article/10.1007/s11098-011-9824-0

Talbott, William 1995: 'Intentional Self-Deception in a Single Coherent Self', *Philosophy and Phenomenological Research*, 55 (1), pp. 27-74.

Thomson, Judith Jarvis 2008: *Normativity*, Chicago: Open Court Publishing Company.

Travis, Charles 2002: 'David Owens, Reason without Freedom', *Ethics*, 113 (1), pp. 173-175.

Vahid, Hamid 2006: 'Aiming at Truth: Doxastic *vs*. Epistemic Goals', *Philosophical Studies*, 131 (2), pp. 303-335.

Velleman, David 2000: *The Possibility of Practical Reason*, Oxford: Clarendon Press.

Walton, Kendall 1978: 'Fearing Fictions', *The Journal of Philosophy*, 75 (1), pp. 5-27.

Wedgwood, Ralph 2002: 'The Aim of Belief', *Philosophical Perspectives*, 16, pp. 267-297.

Whiting, Daniel 2010: 'Should I Believe the Truth?', *dialectica*, 64 (2), pp. 213-224.

Whiting, Daniel 2012: 'Does Belief Aim (Only) at the Truth?', *Pacific Philosophical Quarterly*, http://onlinelibrary.wiley.com/

Williams, Bernard 1973: 'Deciding to Believe', *Problems of the Self*, Cambridge: Cambridge University Press, pp. 136-151.

Williamson, Timothy 2000: *Knowledge and its Limits*, Oxford: Oxford University Press.

Winters, Barbara 1979: 'Believing at Will', *Journal of Philosophy*, 76, pp. 243-256.

Wittgenstein, Ludwig 1969: *On Certainty*, Anscombe, G. and von Wright, G. (eds), Paul, D. and Anscombe, G. (trans.), Oxford: Blackwell.

Wittgenstein, Ludwig 1953: *Philosophical Investigations*, Anscombe, G. and Rhees, R. (eds), Anscombe, G. (trans.), Oxford: Blackwell.

Wright, Crispin, Smith, Barry, Macdonald, Cynthia (eds) 1998: *Knowing Our Own Minds*, Oxford: Clarendon Press.