

Belief, Truth, and Biological Function

Ema Sullivan-Bissett

PhD

University of York

Philosophy

April 2014

ABSTRACT

In this thesis I argue that there is a link between belief and truth, which is indicated by three features of belief relating to how they are formed. I suggest that these features require an explanation from the belief theorist. The first feature is that when we think about what to believe, we find that we immediately and inescapably attend to what is the case (*Transparency*). The second feature is that beliefs are not the kinds of thing which are able to be willed, that is, we cannot *just* believe something, like we can *just* raise our arms (*the Uncontrollability Thesis*). The third feature is that beliefs can be correct (when they are true) and incorrect (when they are false), and are also appropriate targets for claims that they are rational, irrational, justified, unjustified, and so on (*Epistemic Normativity*).

These features, which indicate a link between belief and truth, have been thought to be necessary features of belief, that is, features of *all* beliefs, of *all* believers, across *all* possible worlds. This is a claim which I deny. I argue that accounts which have understood the features in this way cannot explain them, and so ought to be rejected. I offer an alternative account of these features by appeal to the biological functions of our mechanisms for belief-production. My account offers explanations of these features which cast them as *contingent* features of belief, grounded in the biological histories of *some believers*. Belief then is not to be understood as an attitude which is *necessarily* linked to truth.

CONTENTS

ABSTRACT	i
ACKNOWLEDGMENTS	iv
DECLARATION	v
0. INTRODUCTION	1
1: THE MOTIVATIONAL ACCOUNT OF BELIEF	5
1.1 The Motivational Role of Belief	5
1.2 Problems with the Motivational Role of Belief	7
1.3 Belief's Connection to Truth	12
1.3.1 Transparency	13
1.3.2 The Uncontrollability Thesis	18
1.3.3 Epistemic Normativity	27
1.4 The Motivational Role of Belief and Belief's Connection to Truth	30
1.5 Conclusion	34
2: THE TELEOLOGICAL ACCOUNT OF BELIEF	35
2.1 The Truth Aim Teleological Account	36
2.2 Explanatory Work	37
2.3 The Exclusivity Objection	42
2.3.1 Responses to the Exclusivity Objection	44
2.3.1.1 Accepting Exclusivity	44
2.3.1.2 Denying Exclusivity	47
2.4. The Teleologist's Dilemma	53
2.4.1 Response to the Teleologist's Dilemma	55
2.5 Conclusion	60
3: THE NORMATIVE ACCOUNT OF BELIEF	61
3.1 The Normative Account	62
3.2 The Too Demanding Objection	64
3.2.1 Believe Only True Things	66
3.2.2 Propositions We Consider	67
3.2.3 Permissiveness	69
3.3 Explanatory Work	71
3.4 The No Guidance Objection	78
3.4.1 The Norm of Belief as an Idealization	80
3.4.2 Judgement that p	82
3.4.3 Permissiveness and Guidance	83
3.5 Conclusion	85

4: THE ETIOLOGICAL ACCOUNT OF BIOLOGICAL FUNCTION	87
4.1 Historical Accounts and the Connection Between Belief and Truth	87
4.2 The Problem of Function	88
4.3 Etiological Accounts	90
4.4 The Impossibility of Malfunction Objection	93
4.5 Developing the Etiological Account	99
4.6 Responding to the Malfunction Objection	104
4.7 The Circularity Objection	109
4.8 Conclusion	111
5: A BIOLOGICAL ACCOUNT OF BELIEF	113
5.1 Functional Terms	113
5.2 Function: Production of True Beliefs	117
5.2.1 (F1) Our Mechanisms for Belief-Production were Selected for	118
5.2.2 (F2) Our Mechanisms for Belief-Production were Selected for Producing True Beliefs: Pre-empting Worries.....	120
5.2.2.1 (C1) True Beliefs are Not Always Adaptive.....	122
5.2.2.2 (C2) Selection of Mechanisms which Produce True Beliefs is Not the Optimal Adaptive Strategy.....	125
5.2.3 Our Mechanisms for Belief-Production were Selected for Producing True Beliefs	127
5.3 Function: Explanatory Work	135
5.3.1 Transparency	136
5.3.2 The Uncontrollability Thesis	139
5.4 Objections	140
5.4.1 Explanatory Redundancy	140
5.4.2 Contingency is Unacceptable	143
5.5 Conclusion	144
6: USEFUL BELIEF AND EPISTEMIC NORMATIVITY	145
6.1 Truth is Not the Only End	145
6.2 Epistemic Normativity	151
6.2.1 Strong and Weak Epistemic Normativity	152
6.3 Queerness	155
6.4 An Evolutionary Debunking Argument for Strong Epistemic Beliefs	158
6.4.1 Causal Premise	159
6.4.2 Epistemic Premise	164
6.5 An Error Theory for Strong Epistemic Beliefs	165
6.6 Conclusion	168
7. CONCLUSION	171
REFERENCES	173

ACKNOWLEDGMENTS

I dedicate my PhD thesis to my grandmother, Ema Fields, it has always been an honour to be her ‘namesake and champion’, and the vehicle through which she lived her dream.

I cannot thank Paul Noordhof enough, he has been a *really excellent* supervisor, and I feel lucky to have worked with him. He went above and beyond at every turn, displaying brilliant philosophical insight and admirable patience. The work here was immeasurably improved by his unrelenting substantial objections.

I thank Tom Stoneham for his comments on my work, for a number of valuable opportunities during my time at York, and also for starting my interest in the philosophy of delusional belief.

A huge thanks is owed to my friend Jane Tomlinson, who read the whole thesis with an extremely critical eye, and provided (amusingly brutal) very helpful comments, as well as continuous encouragement.

Thank you to the Mind and Reason research group at York, who heard a lot of the work presented here over four years. In particular, thanks to Keith Allen whose comments were always helpful and insightful. Thank you also to Anna Ichino, Barry Lee, Chris Jay, Christian Piller, and Will McNeill, all of whom played an important role in the development of my ideas.

I thank my family for their continuing support, especially Michael Bishop, Glen Sullivan-Bissett, Katie Sullivan-Bissett, Tracey Sullivan-Bissett, Fred Wilson, and Jean Wilson.

Thank you to Rob Davies for reading the work and being supportive in the five days which led up to submission.

I thank my close friends Helen Bradley, Chloe Cordell, Suki Finn, and Nicola Williams for their encouragement and friendship.

Finally, thank you to the Arts and Humanities Research Council and the Royal Institute of Philosophy for funding my doctoral studies.

DECLARATION

I declare that the work presented herein is my own, and that due credit has been given where reference has been made to the work of others. The work presented in §2.3.1.1 was written with Paul Noordhof, and a version of it appears in a paper entitled ‘A Defence of Owens’ Exclusivity Objection to Beliefs Having Aims’ published in *Philosophical Studies*, 2013, Vol. 167, no. 2, pp. 453–7.

Central to any discussion of belief is the question of
the relation between belief and the truth.

(Mayo 1964: 139)

0

INTRODUCTION

It is a common thought that beliefs have a necessary connection to truth, not in the sense that they cannot be false, but rather that it is part of their nature that they are supposed to be true. I will be concerned to give an account of three features of belief which indicate a link between belief and truth. Many philosophers have taken such features to highlight something necessary about the nature of belief. I will deny this. Instead, I seek to answer the challenge that the three features pose without casting such features as illuminating anything necessary about belief, and take the role that truth plays in the fixation of belief as a contingent one. The link to truth had by belief then is not necessary to *belief*, but rather holds contingently of *some believers*.

The thesis is in six parts. In chapter one I will argue that what is essential to belief is its motivational role, and I will defend this view from an objection which claims that such a role cannot be essential to (only) belief, because other attitudes share the role, and so an appeal to belief's motivational role cannot serve to demarcate beliefs from other attitudes. I will then outline the three features of belief which indicate a link between belief and truth. The first is our deliberative belief formation being characterized by Transparency which is the fact that 'when asking oneself *whether to believe that p*' one must 'immediately recognize that this question is settled by, and only by, answering the question *whether p is true*' (Shah 2003: 447). The second is our inability to believe at will, captured by the Uncontrollability Thesis which states that 'unmediated conscious belief-production is impossible' (Noordhof 2001: 248). The third is Epistemic Normativity understood as the claims that belief has a standard of correctness (true beliefs are correctness, false beliefs are incorrect), and beliefs are governed by categorical epistemic norms. Having outlined the explanatory task of the thesis, I will argue that belief's having a motivational role essential to it cannot do the explanatory work called for by the link between belief and truth indicated by these three features, and so something else is required. The motivational role of belief defended in this chapter will form part of the account I develop in chapters five and six, by providing a necessary and sufficient condition on belief which holds for all actual and possible beliefs.

In the next two chapters I will discuss two accounts which have sought to explain the features which indicate a link between belief and truth. In chapter two I will discuss the truth aim teleological account of belief. I argue that the explanations given by the account for the link between belief and truth do not work because something's having an aim is compatible (indeed,

requires), that aim being able to be weighed against other aims, but if this is possible for the aim of belief, the link between belief and truth cannot be explained by appeal to it. This discussion will show that the link to truth should not be understood as located at the intentional level, with truth being something at which the subject is *aimed*, but rather should rather be understood as secured by biology. I also discuss two objections to the teleological account. The first is that our notion of aims is such that aims are the kinds of thing which can be weighed against one another—the so-called ‘aim’ of belief does not allow for this, and so this speaks against belief having an aim. The second is that regulation by the aim does not look common for all beliefs. I will argue that there is not a basic level of truth regulation for all beliefs and so it is not by an appeal to an aim that we can demarcate belief from other attitudes. I will conclude that the teleological account is not able to explain the link between belief and truth, nor have objections to it been answered. The teleological account therefore ought to be rejected.

In chapter three I will discuss the normative account of belief. I will argue that the most plausible version of the normative account is one which casts the norm governing belief as involving permissibility to believe truly, and not obligation to believe truly. It is this version of the account which represents an alternative to my view in so far as it appeals to *sui generis* epistemic normativity in order to explain the three features of belief which indicate a link between belief and truth. I will argue that the account cannot give an explanation of all of these features by appeal to the permissibility norm because this norm is not at the required strength to do the explanatory work—the normativist then, in having to go to a weaker norm, loses her explanation of the link between belief and truth. Next I will consider the No Guidance Objection, which claims that norms are the kinds of things which can guide the activities they govern, the norm claimed for belief is not able to do this, and so this speaks against there being a norm for belief. I will suggest that this objection can be replied to. However, the reply shows that the normative account, like the teleological account, is problematic insofar as it offers an explanation of the link between belief and truth at the intentional level, with the subject exercising the concept of belief, rather than at the sub-intentional level secured by biology. I will conclude that the normative account is not able to explain the link between belief and truth, and so ought to be rejected.

In chapter four I will defend the etiological approach to biological function which I will adopt when formulating my account of belief. I will outline the etiological account before raising an objection which suggests that the account cannot ground a notion of biological normativity, or account for malfunction. This represents a serious concern for my account of belief because it suggests that the normativity of belief can not be reduced to biological normativity. In

response I will offer a development of the etiological account by introducing the requirement of a structural property for functional kind membership. This addition to the etiological account allows for a standard to be laid down by biology, from which token traits can deviate, and hence grounds a notion of biological normativity and allows for malfunction. I also argue that accepting my proposal enables the etiologist to respond to an objection which claims that the individuation of functional kinds in the etiological account is circular. I will then take the etiological account forward into chapters five and six in which I will formulate my account of belief and ascribe functions to our mechanisms for belief-production.

In chapters five and six I will argue that an appeal to biological function can play the theoretical role of accounting for the link between belief and truth. The role of truth played in these features thus comes out as a contingent one grounded in our biological history. In chapter five I will put forward the first part of my account. I will introduce some function terms taken from Ruth Millikan's work, but note that they are to be understood as meeting the conditions laid down by my proposal in chapter four. I will then ascribe the relational proper function of producing *true* beliefs to our mechanisms for belief-production. I argue for this claim by arguing that our mechanisms for belief-production were selected for, and more specifically, that they were selected for the production of true beliefs. It is this second claim which warrants the function ascription I offer. I will then go on to explain how this function ascription explains Transparency and the Uncontrollability Thesis. My explanations have these phenomena not as related to the nature of *belief*, but rather contingently characterizing the beliefs of *some believers*, namely those with a particular biological history.

In chapter six I will suggest that we need to recognise a second function proper to our mechanisms for belief-production. I will claim that a second function proper to our mechanisms for belief-production is the production of useful beliefs (not useful as an approximation to truth, but rather useful in assisting the effective functioning of the believer). I will distinguish between Strong and Weak Epistemic Normativity, and argue that Weak Epistemic Normativity can be explained by appeal to the relational proper function of producing true beliefs outlined in chapter five, and that (the illusion of) Strong Epistemic Normativity, can be explained by appeal to the function of producing useful beliefs. I will argue for Strong Epistemic Normativity being false by appeal to three considerations: its falsity follows from Weak Epistemic Normativity being true, considerations of metaphysical queerness, and my Evolutionary Debunking Argument which shows that beliefs based on Strong Epistemic Normativity are unjustified (which pre-empts the argument that they are justified which might provide traction against my account which claims that they are false). My conclusion will be that my account can explain

Epistemic Normativity, the third of the three features identified in chapter one as requiring an explanation.

I will conclude that the link between belief and truth is a contingent one, secured by the biological histories of actual world believers. Alternative accounts make a mistake in projecting the particular circumstances of *some believers* into a claim about *all belief*. The link between belief and truth is fully explained by appeal to the biological functions of the mechanisms which produce beliefs.

THE MOTIVATIONAL ACCOUNT OF BELIEF AND BELIEF'S CONNECTION TO TRUTH

In this chapter I will briefly outline what has been called the ‘standard way’ (van Leeuwen 2009: 229) of characterizing the motivational role of belief, before defending it from the objection that such a role can be played by other attitudes, and thus cannot distinguish belief from those other attitudes. I will then outline the phenomena to be explained in the thesis; the three features of belief which indicate a link between belief and truth. Finally, I will argue that the motivational approach defended here cannot, on its own, do the work required of the belief theorist in accounting for belief's connection to truth, and so it needs to be supplemented.

1.1 The Motivational Role of Belief

Let us distinguish two theses at the outset. We can understand belief as a state which interacts with desires to cause action, or as a state which does this, but additionally, *when* it has true contents, results in *successful* action.

Frank Ramsey (1931), and following him David Armstrong (1973), make claims in line with the first thesis. They claim that beliefs could be thought of as maps. However, this could also be true of thoughts or mere entertainings, they too could be considered maps of the world ‘however wild or inaccurate’, and so the analogy with maps will not serve to demarcate belief from other attitudes. What can be the fact that ‘*beliefs* are maps *by which we steer*’ (Armstrong 1973: 4). They guide action insofar as they are ‘maps of the world in the light of which we are prepared to act’ (Armstrong 1973: 4).

More recently, claims about the motivational role of belief have been more in line with the second kind of thesis, and get a fix on the motivational role of belief by focussing on what happens when beliefs are *true*, given certain desires. We might think of this more recent claim as being not merely that beliefs are maps by which we steer, but as claiming additionally that if we steer by true beliefs, this will result in the satisfaction of our desires. The motivational role of belief has been claimed to be something like the following: ‘beliefs, jointly with desires, cause and rationalize actions that will make the contents of the desires true, if the contents of the beliefs are true’, Neil van Leeuwen refers to this as the ‘standard way’ of characterizing the motivational role of belief (van Leeuwen 2009: 229). This is indeed a very standard view, and found in several other places. For example, Armstrong understands Ramsey's claim that beliefs are maps by which we steer as suggesting that given ‘suitable dominant desires, then the belief-

state will cooperate with the desire so that they are jointly responsible for the subject's acting in a certain way' (Armstrong 1973: 71). This view is also found in J. T. Whyte's work on success semantics in which he claims that when our beliefs are true 'we get what we want by acting on them' (Whyte 1990: 149). We have moved then from beliefs being states which interact with desires to cause action, to beliefs being states which do this, and additionally, when true, result in successful action. The success of our actions then is related to truth in the motivational role claimed for belief.

This notion of belief which relates true beliefs to successful action—with some development—is the target of David J. Velleman's argument that it fails to capture the distinctive character of belief. Velleman takes the motivational approach as offering a sufficiency condition for a state to count as a belief, that is, he construes the motivational theorist as claiming that 'all that's necessary for an attitude to qualify as a belief is that it disposes the subject to behave in ways that would promote the satisfaction of his desires if its content were true' (Velleman 2000: 255). There are good reasons to think that this characterisation is inadequate as it is too easy for a state to meet. As Lucy O'Brien points out, there are other states of a subject which may meet Velleman's construal of the motivational criterion for belief; a subject may have 'perceptions that p , desires that p , or questions as to whether p ' which may dispose him to form beliefs that p , and they may dispose him to behave such that his desires would be satisfied if p were true (O'Brien 2005: 55). These attitudes would meet the condition Velleman offers, as they would 'dispose [the] subject to behave in ways that would promote the satisfaction of the subject's desires if the contents were true' (O'Brien 2005: 55–6). However, these states are not beliefs, and so Velleman's sufficiency condition gives us the wrong result. Given this, O'Brien claims that the motivational theorist ought to say that for a state to be a belief, all that is necessary is that it '*by itself*,¹ and relative to a fixed background of desires, disposes the subject to behave in ways that would promote the satisfaction of his desires if its content were true' (O'Brien 2005: 56). For a cognitive state to be a belief, it is necessary and sufficient that it alone, when true, disposes the subject to behave in ways which would satisfy her desires. This condition on belief distinguishes it from other attitudes, and rules out the cases O'Brien claimed spoke against Velleman's construal. For example, I could imagine that *there is*

1 One might think about the strength of the 'by itself' clause here. If there were, say, neural necessary conditions for the disposition to be instantiated, then the state does not *by itself* dispose the subject to behave in certain ways. O'Brien does not give an analysis of the 'by itself' clause other than introducing it to rule out those cases in which non-belief states play a mediating role in the manifestation of the dispositions. Let us read the 'by itself' clause in a coarse grained way such that neural necessary conditions are not given a role in the analysis (though they will play a role in the causal story of the subject's action).

food in the fridge and I could believe that *there is food in the fridge*. If I have a desire for food, my imagining that *there is food in the fridge* would not dispose me to go to the fridge,² but if I have a belief that *there is food in the fridge*, that would dispose me to behave such that my desire for food would be satisfied, if the belief were true.³

1.2 Problems with the Motivational Role of Belief

I have briefly outlined a standard way of characterizing the motivational role of belief, being careful to include the ‘by itself’ clause suggested by O’Brien. Here I will consider an objection which has been levelled against the motivational approach to belief, and has traction against the revised account as outlined above.

I find the motivational account of belief antecedently plausible. The reason I focus on Velleman’s objection is because he thinks we need to appeal to the link between belief and truth in order to distinguish belief from other attitudes. As I have indicated, I think the link between belief and truth is not a demarcating feature of belief, but rather is rooted in biology and does not illuminate anything necessary about belief. Velleman’s objection suggests that the motivational role specified for belief in the previous section cannot distinguish belief from other propositional attitudes. It is precisely this function—providing a criterion with which we can demarcate belief from other attitudes—which the motivational role has in my account of belief.

The objection is that the motivational role claimed as sufficient for belief is shared by other attitudes, and so cannot serve to demarcate belief from them. Velleman raises this objection and appeals to the motivational role of imagination. Velleman claims that with respect to its motivational role, beliefs are no different from imaginings; a child imagining to be an elephant will be disposed to act in ways such that, if the content of the imagining were true, it would promote the satisfaction of her desires (Velleman 2000: 255–6). The child in this case then would have her actions guided by her imaginative state. This means that the motivational role of

² Unless my imagining that *there is food in the fridge* brings about in me a desire to know whether or not there is food in the fridge. In this case though, presumably, my desire is for this knowledge and the belief which would promote the satisfaction of this desire if true would be the belief that *walking over to the fridge and opening it would enable me to know if there were food inside*.

³ Of course, there needs to be some *ceteris paribus* clauses read in—I would not go to the fridge even if I had a desire for food and the belief that *there is food in the fridge* if I also believed that all of the food was poisonous. When talking about beliefs combining with desires to give rise to action, suitable *ceteris paribus* clauses should be read as implicit.

belief, so construed, cannot demarcate belief from other cognitive states.⁴ Important to Velleman's discussion is his worry that belief–desire explanations of pretending make the child depressingly adult-like, in touch with reality and believing that they are engaging in a bit of play, will thus act on their imagination as if it is belief, and so on (Velleman 2000: 257–8). So there is an intuition in play here that introducing beliefs about what the child is up to gives a depressing reading of imaginative play, and it is for this reason that imaginings can play the role of belief *even with* O'Brien's 'by itself' clause. When we start introducing beliefs because we think they are required in order for imagining to motivate action, we do not respect the intuition that introducing beliefs gives a depressing reading of child play.

The motivational theorist might instead reject this intuition and claim that it is not the imaginings of the child which are playing the motivational role of belief, but rather beliefs (though of course, not the belief that she is an elephant!). So it is not the child's imagining that she is an elephant which gets her limbs into certain positions, rather it is beliefs like *I am playing a game, elephants move in this way*, and so on (see chapter two of Nichols and Stich (2003) for an example of this strategy). It is this kind of explanation which Velleman takes to portray the child as adult-like.

Let us accept the intuition. This means that the motivational theorist must allow that imaginative states can play the motivational role of belief, because belief doing all of the work makes the child adult-like, which is theoretically unappealing. However, if we concede that imaginative states can motivate action, this is not enough to show that the motivational role of belief as specified earlier cannot be a demarcating criterion of belief. In her discussion of Velleman's case, O'Brien claims that in order to do the work Velleman wants of it (to show that imagination shares the motivational role of belief), the example has to demonstrate not merely that imaginings *can* motivate action, but that they can do so *in the same way* that beliefs do (O'Brien 2005: 57). But Velleman's case does not show this. Rather, assuming she were able to come to believe that she, Lucy O'Brien (LOB), were an elephant (that is, assuming her being an elephant is *believable*), then that state would have a different motivational role to the corresponding imagining. In the case of belief she would 'probably be trying to find ways to resign from [her] job, break it to [her] family, buy a new bed and so on' (O'Brien 2005: 59). But, in the case of imagining, she does not find herself doing any of those things, and the same could

⁴ It is this which led Velleman to appeal to the aim of belief in order to demarcate belief from other regarding-as-true states (see Velleman 2000, chapter eleven. Though he later changed his mind and is now a normative theorist (see Shah and Velleman 2005)). I will discuss the teleological and normative accounts in chapters two and three respectively.

be said of the child engaging in imaginative play (the child engaging in imaginative play does not exhibit the kinds of responses one might expect if she really *believed* that she were an elephant).

Paul Noordhof makes some similar remarks in his discussion of Velleman's case. He claims that though there may be some behavioural similarities in the way the child behaves if she imagines she is an elephant and the way she would behave if she believed⁵ that she were an elephant, there clearly would be a difference. If the child *believed* that she were an elephant, would we not expect a little shock at the transformation? At the very least we would suppose that the child would 'seek to live outdoors [and] consume elephant edibles' (Noordhof 2001: 253). The role that imagining can play in motivating action is context-dependent such that there will be cases in which the roles played by imagining that *p* and believing that *p* come apart. Once we enumerate these contexts, we can distinguish belief from imagination (Noordhof 2001: 253).

Notice too that even if we allow that imaginings can play the motivational role of belief in some contexts (though as we have just seen, imagining that one is an elephant is not one of those contexts), there has to be more to the story—mere imagining cannot alone be efficacious in the production of the child's behaviour:

in pretending, the child seems to be motivated not just by imagining being an elephant. A child can perfectly well be day-dreaming, imagining being an elephant wandering through the plains of Africa, without thereby pretending that he is an elephant. In pretending the subject is also motivated to act out his imaginings. (O'Brien 2005: 57–8)

So we need to distinguish those cases in which a child *merely* imagines that she is an elephant (but this does not motivate her to behave as if she were), and those cases of pretence in which her imagining that she is an elephant motivates her to act in line with that imagining. O'Brien appeals to desires suggesting that it 'is plausible to claim that the child's pretending is motivated by a desire to act out his imaginings' (O'Brien 2005: 58). The desire to act out imaginings might be motivated by the child's desire to play the game of make-believe, his desire to actually *be* an elephant, or even 'by the desire that he keeps track of, and so aids, his imaginings' (O'Brien 2005: 58). As O'Brien points out, if we think that this is what is motivating the child's behaviour, we can claim that the child 'in some sense operates with an awareness of a distinction between that which is real and that which is imagined' (O'Brien 2005: 58). However, she also thinks that

⁵ Noordhof contrasts imagining with judging in this discussion, but his remarks can be translated here into talk of belief from talk of judgement.

telling this kind of story does not lead us to treat the child as depressingly adult-like or disengaged in imaginative play. Indeed, she thinks Velleman ‘may not disagree’ with the story as it is put here, the child ‘operates with a real-world desire, the desire leads to behaviour which is imagination led: it is a desire to be governed by the contents of his imaginings rather than his beliefs’ (O’Brien 2005: 58). So we can concede that imagination can play the role of belief, but we add the caveat that this can only occur in *some contexts*, and not *by itself*.⁶

So, if we accept the intuition that explaining the child’s behaviour by appeal to beliefs she has about what she is up to gives an unappealing characterisation of imaginative play, we should instead allow that the child’s imaginings are motivating her actions. However, this is not to say that imaginings play the *same* motivational role as belief, we saw in Velleman’s case that the child’s behaviour would be very different if it were the *belief* that she were an elephant that was guiding her action. Further, even in cases in which imaginings that *p* do guide a subject’s action in the same way that a belief that *p* would, something else is required, and so the imagining that *p* cannot play the role *by itself*, unlike belief and *across contexts*, unlike belief.

A variant of this objection can be constructed by considering another attitude which looks to share the proposed motivational role of belief, what Michael Bratman has called *acceptance in a context* (Bratman 1992). One might wonder about the argumentative gain of a consideration of Bratman’s acceptance in a context cases. A consideration of one of Bratman’s cases here strengthens the motivational theorist’s opponent’s objection—it makes clearer the problem, not by raising a different issue, but by demonstrating the generality of it. Let us consider an example:

[m]y close friend has been accused of a terrible crime, the evidence of his guilt is strong, but my friend insists on his innocence. Despite the evidence of guilt, my close friendship may argue for assuming, in my ordinary practical reasoning and action, that he is innocent of the charge. In making plans for a dinner party, for example, such considerations of loyalty might make it reasonable for me to take his innocence for granted and so not use this issue to preclude inviting him. Yet if I find myself on the jury I may well think that I should not take his innocence for granted in that context for reasons of friendship. (Bratman 1992: 8)

⁶ A worry might be that belief does not motivate action by itself either, and so imaginings requiring something else for action does not speak against imaginings playing the same motivational role of belief. I claimed that the motivational role of belief was that it by itself, and ‘relative to a fixed background of desires’, disposes the subject to act, across contexts. So beliefs guide in concert with desires, they too do not motivate action *by themselves*. The crucial difference though is that the content of the desire with respect to imagining which is required for the imagining to guide action, regards precisely *acting on* that imagining. Whereas the content of a desire to act on a belief that *p* does not refer to that very acting, but rather to some other good.

In this example, the state of acceptance might be said to play the motivational role specified for belief by O'Brien. The agent accepts that her friend is innocent, despite there being strong evidence of his guilt. In accepting this proposition—*my friend is innocent*—the agent proceeds to include her friend in plans for a dinner party (which presumably she would not do were she to believe he were guilty, and have her action guided by this belief). We might think then, that the proposition accepted by the agent—*my friend is innocent*—is such that it '*by itself*, and relative to a fixed background of desires, disposes the subject to behave in ways that would promote the satisfaction of his desires if its contents were true' (O'Brien 2005: 56).

Though once again, the motivational theorist can respond to worries which might be brought out of considering the motivational role of acceptance in a context by pointing out that this role is not played *by itself* and *across contexts*. Rather, it can only be played in conjunction with some other state, the content of which includes the very acting on the acceptance, and only in some contexts. An acceptance does not and cannot '*by itself* and relative to a fixed background of desires, [dispose] the subject to behave in ways that would promote the satisfaction of his desires if its content were true' (O'Brien 2005: 56), *across contexts*.

This is all the motivational theorist needs to say: the motivational role specified for belief, including the *by itself* clause, is one had only by belief, *across contexts*. Imaginings and acceptances can play that role, but not *by themselves* and not *across contexts*, and it is belief's being able to play the role in these ways which can demarcate it from other attitudes. This response is something the motivational theorist ought to say. It is also sufficient for my purposes, as I want to defend the claim that the motivational role of belief is sufficient to distinguish it from other attitudes—the motivational account then, offers us a necessary and sufficient condition for something to be a belief.

I now have a necessary and sufficient condition on attitudes being beliefs, and I take the modal status of this condition to be strong: it is metaphysically⁷ necessary that for a state to be a belief it '*by itself*, and relative to a fixed background of desires, disposes the subject to behave in ways that would promote the satisfaction of his desires if its content were true' (O'Brien 2005: 56), *across contexts*. With this condition in mind, I now move on to describe three features of belief which an account of belief needs to give an explanation of, these features indicate a link between belief and truth. Although, as we shall see later, the third feature actually reveals something else. In my account I argue that only the first two features indicate a link between

⁷ I do not think the claim here should be stated as conceptually necessary, it seems to me that our concept of belief does not include the motivational role I have specified here.

belief and truth, whilst the third requires a different explanation. Until then I will continue to write as though the third feature is one which indicates a link between belief and truth (as this feature is treated as such by the teleological and normative accounts which I will discuss in chapters two and three respectively).

1.3 Belief's Connection to Truth

The connection between belief and truth forms the explanatory burden of the thesis, and so here I will argue that there is such a connection, by discussing three features of belief which indicate it. It is these features which I will discuss accounts of belief with respect to, and give an alternative account of in chapters five and six. As we will see in chapters two and three, many philosophers offering accounts of the link between belief and truth have taken the link to be a constitutive or conceptual one. On these views, it is part of the very nature of belief, that it is linked to truth, it is a necessary condition on something's being a belief, that it exhibits those features, which indicate a link between belief and truth.

I think that the three features of belief outlined below are not constitutively or conceptually necessary ones. Rather, the motivational role⁸ of belief defended above is constitutive of belief, and the features described in the next three sub-sections are contingent, and these can be explained by appeal to the biological functions of our mechanisms for belief-production. The biological function part of my account which will be elucidated in chapters five and six is more like an 'add-on', an independent piece of theory which can be plugged into a motivational account of belief

⁸ I claim that the motivational role specified for belief is constitutive of belief, that is, it is necessary and sufficient for something's being a belief that it plays the role as specified in §1.2. In chapter five I specify a biological function for our mechanisms for belief-production, which, I argue, explains the role of truth in two of the three features of belief outlined in this chapter. Another way of understanding the motivational role of belief would be that it is constitutive of belief that it has the biological function to play some role, so that it is necessary for something's being a belief, that it is produced by mechanisms which are supposed to produce items which play some role. I find this line unappealing because I think that beliefs are wider than biological creatures; that is, I want to leave it open that beliefs can be had by non-biological creatures, but that such beliefs are not guaranteed to share the connection to truth had by beliefs of biological creatures in the actual world.

1.3.1 Transparency

The first feature of belief which indicates a link between belief and truth is the phenomenon of Transparency (hereafter capitalized when used in the sense which follows). Nishi Shah has given the most attention to this phenomenon, and so in order to help elucidate Transparency I will look at his work. Transparency characterizes our deliberative belief formation. What I mean by *deliberative* here is when one is *deliberating*. So deliberative belief formation, is belief formation over which one deliberates. Transparency is the fact that ‘when asking oneself *whether to believe that p*’ one must ‘immediately recognize that this question is settled by, and only by, answering the question *whether p is true*’ (Shah 2003: 447).

It is important to distinguish two theses, and so two characterizations of the phenomenon. The first is that though truth is not an optional end in deliberative belief formation, the truth of *p* is not always *sufficient* for belief that *p*. So though the truth of *p*, by the subject’s lights, is required for a subject to believe that *p*, it does not suffice as grounds for believing that *p*, there might, for example, be other epistemic grounds related to rational or justified belief, or even non-epistemic grounds relating to the practical consequences of believing that *p*.

A stronger claim is that the question *whether to believe that p* collapses into the question *whether p is true*. Here it is not merely that truth is not an optional end in doxastic deliberation, not merely that truth is *relevant* when deliberating over what to believe, it is the much stronger claim that truth that *p* is *sufficient* for belief that *p*. So if a subject resolves *whether p is true*, for instance—she takes *p* to be the case—then she believes it. There is no further question of whether *p* should be one of her beliefs. Shah should be taken to be positing this stronger claim, and more importantly, it is this thesis which is descriptively true. This is what happens in deliberative belief formation.

Philosophers characterizing the phenomenon of Transparency are keen to make clear that the step from the deliberative question *whether to believe that p* to the question *whether p is true* is ‘immediate and not inferential’ (Engel 2007: 198). When we deliberate over whether to believe that *p*, the truth of *p* is ‘not an optional end for first-personal doxastic deliberation, providing an *instrumental* or *extrinsic* reason that an agent may take or leave at will’ (Shah 2003: 447). If it were then the agent would have to make an inference from discovering that *p* is true and determining whether she ought to believe it, perhaps via what Shah calls a ‘bridge premise’ relating to whether or not it is good to have a true belief with respect to *p*. But there is no such inferential step in doxastic deliberation, ‘there is no such gap between the two questions within the first-personal deliberative perspective; the question *whether to believe that p* seems to collapse into the question *whether p is true*’ (Shah 2003: 447). The two questions then—*whether to believe that p* and

whether p is true are inseparable in doxastic deliberation; one cannot take oneself to have answered the former, without also taking oneself to have answered the latter.

Transparency does not show that deliberative belief formation is such that the products thereof are formed with an eye exclusively as to the truth of *p*, rather, Transparency is just the fact that, ‘one cannot deliberatively, and in full awareness, let one’s beliefs be guided by anything but truth’ (Steglich-Petersen 2006: 503). But that is not to say that *in fact* deliberative beliefs cannot be influenced by non-epistemic factors, rather, it is just that the deliberator cannot take them to be so influenced, first-personally, whilst deliberating. Provided that such non-epistemic factors are not acknowledged by the deliberator, Transparency does not rule out their influencing the content or fixation of deliberative beliefs.

Transparency is a feature had only by beliefs. The phenomenon is only present in deliberation over what to *believe*, not in deliberation over what to *imagine*, *suppose*, *hypothesise*, and so on. When I deliberate over whether to *imagine* that *p*, my deliberation is not transparent to the question *whether p is true*. Equally, when I deliberate over whether to *suppose* that *p*, say, for the sake of argument, my deliberation here is also not transparent to the question *whether p is true*. If I discover evidence for *p*, this is immediately and unavoidably relevant only to whether to *believe* that *p*, not to whether to *imagine*, *suppose*, or *hypothesise* that *p*. Shah claims that in its indicating a ‘substantive relation between belief and truth’ Transparency therefore ‘serves to distinguish belief from other propositional attitudes’ (Shah 2003: 448).

The phenomenon of Transparency is only present in *deliberative* belief formation, that is, it is only present ‘in the context of deliberation that is structured by the question *whether to believe that p*’ (Shah 2003: 467). This means that an explanation of Transparency does not need to apply to instances of non-deliberative belief formation, because Transparency is not present in such cases (though there may be a sense in which mechanisms which produce non-deliberative beliefs are transparent to, or regulated by, truth without the awareness of the subject).

Also, for Transparency to be present in deliberation, it is not required that the deliberating subject explicitly pose to herself the question *whether to believe that p*. Just as a subject can deliberate about what to do in a given situation without explicitly posing to herself the question *what should I do?* Rather, the requirement is only that the subject’s thinking ‘manifests some recognition that this is the question’ that she is seeking to answer, for without this manifestation, her stream of thought would not be deliberation about what to believe or what to do, but would be better described as ‘a stretch of directionless cogitation’ (Shah 2003: 466).

The modal status of Transparency is taken to be strong, so it is not merely a ‘quirky feature of human psychology’ (Shah 2003: 447), a contingent feature of human minds, that our

deliberative belief formation is characterised by Transparency. Rather, it is something in the very nature of belief or deliberative belief formation (to be disambiguated below) which demands transparency to truth considerations. Shah for one takes it that Transparency is ‘expressive of a *conceptual* truth about belief’ (Shah 2003: 468). It is not only metaphysically necessary that deliberative belief formation is characterized by Transparency, but rather that ‘[T]ransparency obtains for all thinkers who engage in doxastic deliberation, and therefore the internal connection between judgments of evidence and belief holds of any creature to which we are willing to attribute the full-fledged concept of belief’ (Shah 2003: 457). Some attempts to account for Transparency have utilized this claim—that it is part of the very concept of belief that deliberative belief-formation should be so governed.

We ought to distinguish two theses, both of which look to be supported by Shah (2003). Again, the focus is on Shah here as a discussion of his remarks will help bring out the precise nature of the phenomenon. The two theses we should distinguish are the following (where constitutive for Shah is conceptual):

- (T1) Transparency is constitutive of deliberation over what to believe.
- (T2) Transparency is constitutive of belief.

There are important differences between these theses, and they express an asymmetric dependence. Though the truth of (T1) follows from the truth of (T2), the reverse does not hold. If (T1) is true, and any deliberation over what to believe will be characterised by Transparency, it does not follow that Transparency is a necessary feature of *belief*. This is because not all beliefs are deliberative. So consider a world in which believers have only non-deliberative beliefs. This would be a world in which believers do not deliberate over what to believe, and so we will not have Transparency. This means then that there is a world in which there are believers, but not Transparency. So this means that Transparency being constitutive of deliberation over what to believe—thesis (T1)—does *not* entail that Transparency is constitutive of belief, thesis (T2). However, if (T2) is true, if Transparency is constitutive of belief, (T1) must follow. If (T2) is true, and bearing in mind that Transparency is only a feature of *deliberative* belief, then it will follow that there are no believers without deliberation, our believers with only non-deliberative beliefs are ruled out—they are metaphysically or conceptually impossible. Belief is just, necessarily, the type of attitude which can be deliberated upon. If this is right, from the claim that Transparency is constitutive of belief (T2), Transparency being constitutive of deliberation over what to believe (T1) follows.

Shah makes some remarks in support of both of these theses. In support of (T1) he claims that the ‘seamless shift in focus from belief to truth [...] is something that is demanded by the *nature of first-personal doxastic deliberation*’ (Shah 2003: 47, my emphasis). This looks to be precisely the claim that Transparency is constitutive of deliberation over what to believe. He also refers to the role of truth in structuring deliberation over what to believe as ‘noncontingent’ and ‘noninstrumental’ (Shah 2003: 448), and claims that ‘[T]ransparency is the consciously felt authority of truth for belief in *any deliberation that aims to settle belief*’ (Shah 2003: 468, my emphasis). Finally Shah claims that ‘determining *whether p is true* will be immediately imperative, to the exclusion of any other question, for anyone who entertains the deliberative question *whether to believe that p*’ (Shah 2003: 470).

However, Shah also makes some remarks which support a reading that he is making the claim of thesis (T2), that Transparency is constitutive of belief. We saw earlier that he claims that ‘[T]ransparency is expressive of a conceptual truth about *belief*’ (Shah 2003: 468, my emphasis), and he also claims that ‘[T]ransparency can’t be the conscious face of a merely metaphysical truth about belief, [it] must express a conceptual truth about belief; a truth that an agent grasps merely in virtue of treating his deliberation as answering to the question *whether to believe that p*’ (Shah 2003: 468–9). The quoted material here lends support to both thesis (T1) and (T2), and Shah looks to treat them as amounting to the same thing. But Transparency expressing ‘a conceptual truth about belief’ does not follow from it being ‘a truth that an agent grasps in virtue of treating her deliberation as answering to the question *whether to believe that p*’ (recall our other-world creatures who only have non-deliberative beliefs).

Shah thinks that Transparency is conceptually necessary to belief, so to understand the concept of belief one must understand it as an attitude which answers to truth considerations. He also claims, with Velleman, that what makes something a belief is the way in which it is ‘formed, revised, and extinguished’, for short; ‘the way in which it is regulated’ (Shah and Velleman 2005: 498). So all beliefs—even non-deliberative ones—are regulated for truth, and this too is a conceptual claim about belief.

I think that thesis (T2) is false. If Transparency is constitutive of belief, then the motivational role I specified for belief cannot be sufficient for a state to count as a belief, Transparency must also be present. Shah would welcome this result, with Velleman he claims that ‘the motivational role of belief is one that it shares with other cognitive attitudes’ (Shah and Velleman 2005: 497–8), and so it cannot demarcate beliefs from other attitudes (I argued in §1.2 that this is false). Rather, ‘[T]ransparency indicates a substantive relation between belief and truth, it serves to distinguish belief from other propositional attitudes’ (Shah 2003: 448).

Here are some reasons for thinking that thesis (T2) is false. If true, it rules out solely non-deliberative believers as a metaphysical or conceptual possibility. It would follow from the truth of (T2) that it is at least metaphysically impossible (though Shah would make a conceptual impossibility claim), that there could be *believers* who did not deliberate over what to believe, they only had non-deliberative beliefs. Creatures of this kind are ruled out by thesis (T2), because it is part of the very nature of belief, that it is characterized by Transparency, and Transparency is only present in deliberative belief formation. This gives us reason to think that thesis (T2) is false: if what it takes to be a belief is truth regulation, as we saw above from Shah and Velleman (2005: 498), then we could have believers who had only non-deliberative attitudes which were regulated for truth who would thus count as believers (though they would also be believers who did not possess the concept of belief).⁹ We might think that creatures of this kind exist in the actual world, namely, non-human animals. They look to be believers, and yet they do not *deliberate* over what to believe, and so Transparency is not constitutive of their beliefs.

Shah and Velleman pre-empt a similar worry, when they say that human children and some non-human animals have beliefs even though they do not have the concept of ‘truth conducive methodology’, but this is just to say that such believers are ‘believers unwittingly’; ‘they just can’t conceive of what they have as beliefs’ (Shah and Velleman 2005: 520). So it is a condition on *conceiving* of an attitude as a belief that a subject understands it as an attitude governed by Transparency, and not a condition on an attitude *being* a belief, and so human children and non-human animals can have beliefs, they are just unable ‘to classify their beliefs as such’ (Shah and Velleman 2005: 534, n. 43).¹⁰ So Shah and Velleman think that there can be unwitting believers, believers who do not possess the concept of belief. But this only deals with the worry that possessing the concept of belief as an attitude which is governed by Transparency is necessary to be a believer. The worry I have raised with thesis (T2) is that to be a *believer* one must, at least sometimes, have beliefs formed through doxastic deliberation because Transparency is necessary to belief. So the states a creature has can only be classified as beliefs if they are sometimes brought about deliberately, because Transparency is constitutive of belief, and only characterises deliberative belief formation.

⁹ Davidson claims that a creature cannot have a belief if it does not have the concept of belief (Davidson 2001: 170). If this is right, this would rule out our solely non-deliberative believers who do not exercise the concept of belief when forming beliefs. It would also rule out beliefs had by children and non-human animals. I put this position aside.

¹⁰ We might well ask here why the concept of belief should have this feature if belief itself does not. My preferred view does not face this kind of worry because I make no conceptual claim about belief.

Thesis (T1) looks a more plausible candidate for how to understand Transparency. It follows from thesis (T1) that there is no non-transparent deliberation, deliberation over what to believe is necessarily characterized by Transparency. So there are no other-world believers whose deliberative belief formation is not characterized by Transparency. Though I will deny this later, it does not look as obviously problematic as the claim that there are no other-world solely non-deliberative believers.

What we need an account of then, is why the question *whether p* answers the question *whether to believe that p* in deliberation over what to believe. My account gives an explanation of Transparency which takes it to be a contingent feature of belief (§5.3.1), an explanation of which can be found by looking to one of the biological functions of our mechanisms for belief-production. This means that I allow for other-world believers whose deliberative belief formation is not characterised by transparency to considerations of truth.

1.3.2 The Uncontrollability Thesis

The second feature of belief which indicates a link between belief and truth is our inability to bring about beliefs at will. Our inability to believe at will falls out of deliberative belief formation being governed by Transparency.

There have been several attempts at pinning down precisely what we cannot do—precisely what this inability amounts to—so I will try to bring together what philosophers have agreed on regarding this, in order to show consensus. In the end though, I will take on Noordhof's Uncontrollability Thesis as that which requires an explanation from the belief theorist.

Bernard Williams claimed that 'I cannot bring it about, just like that, that I believe something' (Williams 1970: 148), while Keith Frankish has claimed that 'most writers deny that we can form beliefs *directly*, in a causally unmediated way, simply by forming and executing an intention to do so' (Frankish 2007: 523). Williams and Frankish have in common the clause that for a case to count as willed belief, there can be no causal mediation between the intention to believe something, and the believing, Frankish makes this point explicit, whilst we can draw it out from Williams's phrase 'just like that'.

Other philosophers have set out what the inability amounts to more formally. Barbara Winters sets three necessary conditions which must be met for a case to count as a case of believing at will. The first condition is that:

the belief must have been acquired directly and as a result of intending to hold it. (Winters 1979: 244)

Winters claims that what this condition rules out is a case in which I acquire a belief by concentrating really hard on particular evidence or discrediting evidence which stands against the belief I am trying to form. Presumably though, it also rules out cases of hypnosis or pill-taking—so the case where I want to form the belief that *p*, and in order to do so, I am hypnotised to believe that *p*, or I take a pill which induces in me the belief that *p*. This first condition rules this out by requiring that the belief is acquired *directly*, that is, not via a motivated biasing of evidence, or hypnosis, or pill-taking, but also ‘as a result of intending to hold it’. This last clause—‘as a result of intending to hold it’—is also important, because it rules out cases in which I form a belief that *p directly* as a result of *perceiving that p*, or *being told that p*. Winters’s first condition then rules out cases of mediated belief formation, and beliefs not resulting from an intention to hold that belief.

Winters’s second condition on belief at will is that:

the belief be acquired independently of any consideration about its truth. (Winters 1979: 244)

Winters has this condition in order to rule out cases where the subject considers data relevant to the truth or falsity of *p* in her coming to believe that *p*. Winters uses ‘data’ here instead of ‘evidence’ as she has something wider in mind than ‘evidence’ can capture, she wants to include things like testimony—such data is referred to collectively as ‘truth considerations’. Of course, many writers would be happy to take testimony that *p* to be evidence for *p*, so we might instead say that this condition is in place to rule out cases where the subject considers *evidence* for the truth or falsity of *p* in her coming to believe that *p*.

The third condition Winters places on willed belief is that:

[t]he action of acquiring the belief at will be performed with the agent fully aware that he or she is attempting to arrive at the belief this way. (Winters 1979: 245)

Winters includes this condition to rule out ‘[i]rrelevant difficulties arising from suppressed awareness’ (Winters 1979: 245). I do not see the motivation for this third condition. If it is there to rule out cases in which the subject has suppressed awareness of the way she came to hold the belief that *p*, because she came to hold it deviantly, through hypnosis or pill-taking, then this case is ruled out by the first condition. I suspect that Winters has something different in mind. I think she wants to rule out cases in which the subject forms the belief that *p*, directly, and as a result of intending to believe that *p*, and acquires this belief independently of any truth considerations, but whilst she is doing this, she is not fully aware that she is trying to get to

holding the belief in this way. She must then, whilst forming the belief, ‘realise that truth considerations play no role in the attempt’ (Winters 1979: 245). I do not know what this kind of case would look like, if the subject intends to believe that p , and she comes to believe that p directly as a result of this intention, and not via awareness of, or attention to, truth considerations, it is difficult to see how she could fail to realise that considerations of truth played no role in the belief formation. Perhaps what Winters has in mind are cases of *wilful self-deception*, in which subjects want to have the belief that p for instrumental reasons, as a means to some end, rather than because they want the belief to be true (van Leeuwen 2007a: 424). If self-deception were the result of an intention to deceive oneself, then perhaps such cases would meet conditions one and two of Winters’s conditions on belief at will, but importantly, would fail to meet condition three (on most accounts of self-deception), and it is in their so failing, that they do not count as willed believings. I do not think self-deception involves an intention to deceive oneself, but I cannot argue for that here. Rather, I will end my outlining of Winters’s account of belief at will by suggesting that the motivation for the third condition is not quite clear.

The various requirements for a case to count as willed belief, found in the work of the philosophers so far discussed, are captured by Noordhof’s Uncontrollability Thesis, which is the claim that ‘unmediated conscious belief-production is impossible’ (Noordhof 2001: 248). He claims that what is not possible is the conjunction of the following:

- i. I form the conscious intention to believe that p now;
- ii. as a result of having this intention, I believe that p now without there being any mediating act which helps to produce this belief or the support of evidence for p or a belief that by believing that p I make p true;
- iii. the intention and the belief stand in the same relation to each other as intention and action do in the case of intentional actions. (Noordhof 2001: 248)

Like Winters, Noordhof requires that the subject intends to believe that p , and this requirement constitutes his first condition. Though he is more careful than others in that he makes more specific the nature of the required intention, by requiring that it be conscious. Requiring that the intention to believe that p be conscious rules out cases of self-deception where self-deception is characterised as intentional belief formation of a motivationally biased belief. If one takes an intentionalist view of self-deception—which, as I said earlier in relation to Winters, I do not—then in a case of self-deception so construed, the subject could be characterised as (unconsciously) intending to believe that p , and as meeting (ii) and (iii) of Noordhof’s

formulation. The addition of ‘conscious’ here in characterising the intention, rules out intentionalists about self-deception counting cases of self-deception as cases of willed belief, or at least, these cases might count as cases of willed belief but not the kind of case of willed belief which is impossible.

Noordhof’s second condition has much in common with Winters’s first. Here he wants to trace the cause of the belief that p to the intention to believe that p . This looks like Winters’s first condition, that ‘the belief must have been acquired directly and as a result of intending to hold it’ (Winters 1979: 244). As we saw, what Winters’s condition rules out are cases in which a subject could acquire a belief by concentrating on particular evidence, or by being hypnotised, or taking a pill. Noordhof’s second condition is much more explicit to this effect: he rules out the possibility of any ‘mediating act’ which could help produce the belief, this might be the act of getting oneself hypnotised, or taking a pill. He also explicitly rules out coming to believe that p by considering evidence for p , a move Winters also makes in her second condition which requires that the agent does not take into account any truth considerations. Finally, Noordhof rules out cases in which by forming a belief that p , one makes p true, an example of such a case would be something like the following: I offer to give you £100 tomorrow if you can believe now, that tomorrow you will be £100 richer. Here, by forming the belief that you will be £100 richer tomorrow, you make it true that you will be £100 richer tomorrow. In adding in this clause to condition (ii), Noordhof rules out this kind of case as counting as a case of belief at will, or again, at least, this kind of case is ruled out as the kind of case which is claimed to be impossible.

Noordhof’s final condition requires that the intention to believe that p , and the resulting belief that p , are related as intention to φ and φ -ing are in cases of ordinary intentional action. Noordhof claims that this is needed to rule out cases in which ‘a deviant causal chain connects my intention with my belief’, in which case, we would not have a case of willed belief (Noordhof 2001: 248).

I will follow Noordhof in endorsing the Uncontrollability Thesis, that ‘unmediated conscious belief-production is impossible’ (Noordhof 2001: 248). It is this which requires an explanation, and which constitutes the second feature of belief which indicates a link between belief and truth. On my account, the explanation for why the Uncontrollability Thesis holds falls out of an explanation of why Transparency holds (though that is not to say that the only explanation of the Uncontrollability Thesis must go via an explanation of Transparency, see §3.3).

Many philosophers have taken our inability to believe at will as necessary. Noordhof takes it this way, claiming that ‘there is plausibility in the claim that the Uncontrollability Thesis should be interpreted as claiming that (i) to (iii) are jointly metaphysically impossible’ (Noordhof 2001: 248–9). Noordhof is not alone in taking our inability to believe at will in this sense, indeed, ‘there is a widespread sense’ that ‘there is something in the nature of belief that makes it impossible to decide to believe a proposition for which one lacks epistemic support’ (Frankish 2007: 528).

I think that the Uncontrollability Thesis has weak modal status, insofar as the impossibility is not a metaphysical impossibility, but only a contingent impossibility. Here I will briefly outline two arguments in support of the claim that we cannot believe at will. Both argue for something like the modally strong version of the Uncontrollability Thesis. I will suggest that these arguments only support a modally weak version of the Uncontrollability Thesis, and that an explanation of the Uncontrollability Thesis can thus cast it as holding contingently, without this being a mark against such an explanation (see also §5.4.2). The first argument comes from Scott-Kakures, he denies that the following belief transition is possible:

[a]t t , an agent does not believe that p and at $t+1$ the agent believes that p , and the transition from the one belief state to the other is accomplished by a direct and unmediated willing to believe that p . (Scott-Kakures 1993: 77–8)

He argues that believing a proposition *at will* is impossible because ‘nothing could count as *willing* a belief’ (Scott-Kakures 1993: 92). This conclusion is grounded in his necessary condition on intentional basic action. He claims that for an event to count as an intentional basic action it must be produced and guided by some contentful mental state or states. It is in virtue of such guidance that such an event can be said to be *monitored* (Scott-Kakures 1993: 92). When I reach out for a glass, I do so guided by my desire to reach out for a glass. And it is this, at least in part, which makes it such that my reaching out for a glass is an intentional basic action.

Given this, how would a story about willed belief go? Scott-Kakures claims that willing a belief could not count as an intentional basic action because it could not be guided in the appropriate way. He claims that when I formulate my intention to believe that p , my current cognitive perspective (from which such an intention is generated) also includes two further beliefs. The first is that I do not currently, already believe that p . The second is that none of the beliefs I have give sufficient epistemic justification for believing that p . When I formulate the intention to believe that p then, I ‘must regard my current cognitive perspective as not sanctioning the belief that p ’ (Scott-Kakures 1993: 94). I must recognise, upon formulating the

intention to believe that p , that the cognitive perspective I currently inhabit disallows the formation of this belief.

If this is right, contends Scott-Kakures, then it is not possible for me to ‘see my way through to my altered cognitive perspective at $t+1$ ’ (Scott-Kakures 1993: 95). I cannot see my way from t , where I have the intention to believe that p (and do not currently believe that p), through to $t+1$, where I believe that p (as a result of my willing to). And if this is the case, my intention to believe that p cannot be one which governs my behaviour through to my resulting state at $t+1$ (my believing that p). Consequently, my arrival at the belief that p at $t+1$ is something which is ungoverned, and hence cannot be the result of an intentional action, of my willing to believe that p (Scott-Kakures 1993: 95).

So we have an explanation of why we cannot will a belief. Scott-Kakures does two things in order to reach his conclusion. First he offers a necessary condition on intentional basic action. Second he argues that a willed belief could not meet this condition. I cannot will a belief that p , because when I form the intention to do so, I would have to recognise that my current cognitive perspective would not sanction a belief that p . And this is because of two beliefs I already hold: *I do not currently believe that p* and *nothing I do believe epistemically justifies believing that p* . Given this, I cannot get myself from not believing p at t to believing p at $t+1$. I cannot, then, will a belief. ***

However, Scott-Kakures’s argument only demonstrates the contingent impossibility of belief at will. The necessary condition on intentional basic action is such that willed belief cannot meet it. I cannot will a belief, because I recognise that my cognitive perspective does not sanction it. So far this is in line with the Uncontrollability Thesis expressing a contingent claim. The reason my cognitive perspective does not sanction the belief that p , is because of two beliefs I already hold: *I do not currently believe that p* and *nothing I do believe epistemically justifies believing that p* . I take it that this second belief which does not license the belief that p does not do so only contingently, or at least, there is nothing in Scott-Kakures’s argument to suggest anything stronger. So we can allow Scott-Kakures his condition on intentional basic action, but we do not need to allow that belief is *necessarily* monitored by the states it is monitored by in the actual world. Consider other-world believers who believe that p if and only if they believe that believing that p will make them happy. Such a believer may produce a belief even when their current cognitive perspective includes the beliefs *I do not currently believe that p* and *nothing I do believe epistemically justifies believing that p* . Holding such beliefs would not, for these creatures, prevent a belief from being formed at will.

So it is because our belief formation is tied up to truth considerations (or *transparent* to truth considerations) that the belief that nothing I believe epistemically justifies believing that p

means that I cannot will a belief that p . But there is nothing in Scott-Kakures's argument which precludes taking transparency to truth considerations to be a contingent fact about belief. If Transparency were a conceptual or metaphysically necessary feature of belief, then Scott-Kakures's argument would show that belief at will is metaphysically impossible, but I deny this. All that is necessary (and sufficient) for something's being a belief is its playing the appropriate motivational role.

To assume that there is only one form of monitoring that might hold for belief is to assume what needs to be proven. In the actual world, monitoring will be by epistemic considerations because the connection with truth holds. But monitoring might take place in a different way for other-world creatures, for whom the connection between belief and truth does not hold. They see through to the willed belief not via epistemic reasons, but via reasons relating to happiness. So in the case of the other-world believers, for whom biology—nor anything else—secures the connection to truth, the question of something not being sanctioned by our epistemic perspective does not have the same weight. In these circumstances the creature can produce a belief because it will make her happy just like she can raise her arm because it will make her happy. So willed belief is only impossible if the epistemic perspective arises, but it does not for those creatures whose beliefs are not linked with truth. So the epistemic point is irrelevant and the monitoring which goes on for our other-world believers is of the same kind as the monitoring which goes on when I raise my arm. For these creatures, monitoring being secured by considerations relating to happiness does not speak against the product of that monitoring (a belief) being willed, because this kind of monitoring is in play when I (wilfully) raise my arm because doing so will make me happy. Such creatures represent a counterexample to the Uncontrollability Thesis understood as necessarily true on the basis of Scott-Kakures's argument.

I move now to Keith Frankish's argument for the necessary impossibility of belief at will. Frankish recognises that despite consensus, 'it has been hard to articulate a sound argument for this conclusion' (Frankish 2007: 528). Frankish introduces three terms in his argument. First, he calls propositions 'unsupported' if they are ones for which we do not have epistemic support (Frankish 2007: 531). Second, 'deviant' beliefs are those which are produced by desire or other non-truth related influences (Frankish 2007: 531). Finally, when one forms a mental state with the content p , in full consciousness, without any reason to think that p is true, and for pragmatic reasons, the act of forming that state is 'wanton' (Frankish 2007: 533).

Frankish also introduces two principles. The first is the Revised Williams's Principle¹¹ which has it that:

[f]or any proposition *p*, it is impossible to believe in full consciousness both that one consciously believes that *p* and that one's belief that *p* is both unsupported and deviant. (Frankish 2007: 532)

The second is the Control Principle which states that:

intention formation presupposes control; one cannot intend to perform a certain action if one believes that, in normal conditions, one has no control over whether or not one would succeed in performing it. (Frankish 2007: 534)

The argument for the impossibility of believing at will proceeds as follows: if one formed a mental state *M* with content *p* wantonly (consciously, without reason to think *p* true and for pragmatic reasons), one would thereby know that the state which resulted was unsupported (lacked epistemic support) and deviant (produced by non-truth related considerations). Given that one knows this about *M*, one would not be able to regard it as a belief (Frankish 2007: 533). If this is right, the exercise of believing at will becomes self-defeating—one cannot believe at will because attempts to do so would result in products which 'would be unable to survive reflection on their status' (Frankish 2007: 533).

So, argues Frankish, in order to believe at will one of two things must happen. Either one must (alongside forming the belief that *p*), also form the belief that the belief that *p* is supported. Or, one must forget that the belief that *p* has been formed on pragmatic, non-evidential grounds (Frankish 2007: 533). Both of these things are not within our control. Consequently, in order for one to be successful in one's attempt to produce a belief at will, one must also bring about conditions which are not under one's control (Frankish 2007: 534). Clearly, this is an impossible feat. Frankish takes it that this impossibility is 'not contingent but follows from more basic principles, namely the Revised Williams's Principle and the Control Principle' (Frankish 2007: 536).

Frankish's argument goes some way to explaining why the Uncontrollability Thesis holds but the impossibility cited is not shown to be anything stronger than a contingent one. If we

¹¹ Frankish calls the following the Williams's Principle: 'for any proposition *p*, it is impossible to believe in full consciousness both that one believes that *p* and that one's belief that *p* was formed at will' (Frankish 2007: 529). He raises three problems with this principle, and so revises it to meet them.

formed a mental state M with content p *wantonly*, by which Frankish means without any reason to think p is true and for pragmatic reasons, we would know that this state was *unsupported*, that is, lacked epistemic support, and *deviant*, produced by non-truth related considerations, and given that we know this, we could not regard the resulting state as a belief, and so the exercise of belief at will is self-defeating.

If we give ‘unsupported’ and ‘deviant’ epistemic readings, I see no reason to think that the Revised Williams’s Principle is anything but a contingent truth. The definitions Frankish gives of ‘unsupported’ and ‘deviant’ are epistemic ones, so propositions are *unsupported* if they are ones for which we do not have epistemic support, and propositions are *deviant* if they are produced by non-truth related influences (Frankish 2007: 531). Reading the Revised Williams’s Principle with these particular definitions in place suggests that the principle is contingent. It is certainly the case that *we* could not consider as a belief a propositional attitude which was one for which we did not have epistemic support and one which was produced by non-truth-related influences. Also, when *we* attempted to produce a belief at will, *we* would know that the resulting state would indeed be unsupported and deviant in this way. And knowing this, *we* could not consider that state a belief, but why think that this is something to do with the nature of *belief*, rather than the nature of *some believers*? Think back to our other-world believers who believed that p only if they believed that believing that p would make them happy. For these believers, the Revised Williams’s Principle (with epistemic readings of ‘unsupported’ and ‘deviant’) is false—they could consider a state a belief even if it was not epistemically supported and was produced by non-truth-related influences.

However, if we allow *unsupported* and *deviant* non-epistemic readings, then the Revised Williams’s Principle is necessarily true. So for our creature who believes only what she believes will make her happy: if she formed a mental state with content p *wantonly* (consciously, without reason to think believing p would make it happy), she would know that the state which resulted was *unsupported* (was not something likely to make her happy) and *deviant* (produced by non-happiness related considerations). Given that she knows this about M , she would not be able to regard it as a belief, and so the exercise of believing at will would be self-defeating. So if we give *unsupported* and *deviant* readings to reflect the considerations in play for different kinds of believers, then Revised Williams’s Principle is necessarily true. What is contingently true is the role of truth played in the Revised Williams’s Principle.

However, even if the Revised Williams’s Principle is necessarily true, it does not follow that the Uncontrollability Thesis expresses a necessary truth. This is because Frankish has appealed to deliberation in order to explain why we cannot believe at will. He requires that the subject

recognises her beliefs are *unsupported*, which would require deliberation on their epistemic status. However, notice that it is not written into the Uncontrollability Thesis that a belief must be produced *deliberatively*, only *consciously*. And conscious belief-production does not require *deliberative* belief-production. Willed belief does not require deliberation. Just as I can *consciously* and *wilfully* raise my arm without deliberation—without deliberating on *whether to raise my arm*—so too can our other-world creatures *consciously* and *wilfully* produce belief, nothing in Frankish’s argument rules this out. It is not part of the Uncontrollability Thesis that the subject must view her belief as *unsupported* (recall that I rejected Winters’s third condition on belief at will which made this kind of claim).

My account gives an explanation of the Uncontrollability Thesis which takes it to be a contingent feature of belief, an explanation of which can be found by looking to one of the biological functions of our mechanisms for belief-formation. This means that I allow for other-world believers who can believe at will. What requires an explanation from belief theorists then is not the Uncontrollability Thesis understood as necessarily true, but rather understood as contingently true. We will see in the next two chapters that the teleological and normative accounts offer an explanation for a modally stronger version of the Uncontrollability Thesis, I have shown here that we need not take the thesis to be true at this strength.

1.3.3 Epistemic Normativity

The third feature of belief which indicates a link between belief and truth is Epistemic Normativity (though see the end of §1.2 for a caveat on this), which I will understand as the conjunction of two claims. Firstly, belief has a standard of correctness,¹² according to which true beliefs are correct and false beliefs are incorrect. It ‘is part of the “job description” of belief as a distinctive propositional attitude that beliefs are correct or incorrect depending upon the state of the world’ (Railton 1994: 74). Whilst other cognitive states can have contents which are true or false, truth and falsehood are a ‘dimension of assessment of beliefs as opposed to many other psychological states or dispositions’ (Williams 1970: 136). I could have an imagining or supposing with a true or false content which is nevertheless not appropriately judged as correct or incorrect. Correctness conditions then follow not only from the propositional content of a state, but also from the state itself. So beliefs which are correct, are beliefs which are true, my ‘belief that snow is white is correct just in case the belief is true, just in case snow is white’

¹² It is important not to take belief’s standard of correctness as a demarcating feature of belief, as this standard is shared by guesses (see Owens 2003).

(Gibbard 2005: 338). And though we evaluate beliefs with respect to values other than truth (their being rational, justified, and so on), ‘the end of realizing the truth enjoys a special sort of status when it comes to the evaluation of belief’ (Grimm 2009: 256). Put for more formally we can capture the claim that belief has a standard of correctness with the claim that ‘[f]or any p : the belief that p is correct if and only if p is true’ (Bykvist and Hattiangadi 2007: 277).

The standard of correctness for belief is understood as the claim that a belief is correct if and only if it is true. It is important to point out why this is not a trivial claim. If ‘correct’ is taken to be synonymous with ‘true’, the claim that true beliefs are correct beliefs would be trivial. However, a belief’s being correct is not the same as a belief’s being true. Rather, if we say that a belief is *correct*, we are saying that in having that belief, ‘one has got things “right”’, one’s belief is ‘appropriate’ (Wedgwood 2002: 267). Equally, if a belief is *incorrect*, the believer has made a mistake. ‘Correct’ then is not synonymous with ‘true’, but is a normative notion which attaches not to the proposition believed, but to the attitude or act of believing.

Further, if ‘correct’ were synonymous with ‘true’, we would be able to say of other cognitive states that they were *correct* in virtue of having true contents. But this is not something we can do—an imagining with a true content is not therefore a *correct* imagining, nor is a hoping or desiring with a true content a *correct* hoping or desiring. Rather:

‘[c]orrect’ means something distinct from ‘true’, because it refers to the existence of a standard or norm, and implies that the attitude in question conforms to it, and that the content of the attitude is true *in virtue* of the way it is formed, or acquired, or entertained. (Engel 2013: 202)

The other feature of Epistemic Normativity I am interested in giving an account of is the norms which are claimed to govern belief formation. The most obvious and discussed of such norms is grounded in belief’s standard of correctness, Pascal Engel claims that it ‘is generally agreed that the normative dimension in belief is its dimension of correctness, and that the norm for belief, if there is such a norm, is that a belief is correct if and only if it is true’ (Engel 2007: 180).

Other kinds of norms I am interested in giving an account of include norms of ‘evidence, (a belief is correct if it rests upon sufficient evidence), knowledge (a belief is correct if and only if it aims at knowledge), rational norms (a belief is correct if and only if it is rational)’ (Engel 2007: 181). These norms govern only belief—it is inappropriate to say of my imaginings or supposings that they are rational, irrational, justified, unjustified, and so on. There will be exceptions here. One might think that judgements of rationality are applicable to some imaginings. For example, I find myself imagining that p , which frightens me. Nevertheless, I

continue to engage in the imaginative process, that is, I continue to imagine that p . Perhaps it would be appropriate to say of me here that I am acting irrationally or that my imagining that p is irrational. The difference between cases such as these and cases of belief is that the judgements of rationality are grounded in what the act or attitude thereby brings about, and not merely on the contents of the state in question. My continued imagining that p is irrational because it frightens me, I know that it frightens me, and I continue to engage in the act. However, in the case of belief judgements about their rationality are not (or at least not always) concerned with what those beliefs bring about. I may hold an irrational belief that there are seventeen gremlins living on Pluto. Let us suppose that this is a belief upon which I would never, nor need ever, act. It is nevertheless deemed irrational or unjustified, and this is appropriate. Imaginings have different conditions for the application of judgements of justification and appropriateness.

These epistemic norms have also predominantly been understood as categorical ones, that is, they are norms which ‘apply to agents categorically; their reason-giving force transcends agents’ desires, ends, activities, or institutions’, as opposed to hypothetical norms whose ‘reason-giving force depends on agents’ desires or ends, their engagement in certain rule-governed activities, or their occupation of certain roles’ (Olson 2011: 80). Our obligations to comply with these norms are those ‘to which the practical benefits of beliefs are not relevant. They are obligations that arise from a purely impartial and disinterested perspective’ (Feldman 1988a: 236).

What I take to require an explanation from the belief theorist are the following two claims:

- (EN1) Beliefs have a standard of correctness.
- (EN2) Beliefs are governed by categorical epistemic norms.

I will take these two claims to constitute Epistemic Normativity. We might have construed both claims in terms of standards of correctness, or both claims in terms of norms. So instead of (EN2) we might have said that beliefs are governed by standards of correctness with respect to how they are formed, as well as with respect to their content (as captured by (EN1)), and instead of (EN1) we might have said that beliefs are governed by a norm of truth as a way of understanding the standard of correctness. The explanatory burden on the belief theorist is to give an account of why (EN1) and (EN2) hold or at least, why we think that they do (when they do not). My account will pave a middle way, that is, I will argue that there is a sense in which (EN1) and (EN2) are true, but not in the way we take them to be. In chapter six I will argue that

biological norms lay down a standard from which traits can deviate, and so there is a sense in which true beliefs are correct and false beliefs are incorrect (EN1), and a sense in which beliefs are governed by norms (EN2), though such norms are reducible to biology (though still categorical, insofar as they hold regardless of the interests of the subject). Because my explanation of Epistemic Normativity appeals to biological function, I take it to be a contingent feature of belief. Though again, as we saw with Transparency and the Uncontrollability Thesis, the accounts I will discuss in the next two chapters, will take themselves to be explaining a modally stronger feature.

1.4 The Motivational Role of Belief and Belief's Connection to Truth

So far I have outlined a standard characterisation of the motivational role of belief, and defended it from the worry that other attitudes can play that role. I have also outlined three features of belief which require an explanation. These are Transparency, the Uncontrollability Thesis, and Epistemic Normativity. The Uncontrollability Thesis, I suggested, followed from deliberation being characterized by Transparency, and so an explanation of the Uncontrollability Thesis might be derivable from an explanation of Transparency. Finally I understood Epistemic Normativity as requiring that the belief theorist give an account of why (EN1) and (EN2) hold, or at least, why we think that they do.

I now suggest that the motivational approach to belief defended at the start of this chapter is not, by itself, able to do the work required to account for the connection between belief and truth; it cannot explain why Transparency, the Uncontrollability Thesis, or Epistemic Normativity hold, nor is doing so in its remit. Though, as I claim that the motivational role of belief is essential to belief, it is important to point out why this claim is not able to give an account of the connection between belief and truth, and so demonstrate that the motivational account needs supplementing in certain ways in order to account for the three features of belief I have outlined.

What could a motivational theorist say about these features? I characterised the motivational role of belief earlier by following O'Brien: a state is a belief if it '*by itself, and relative to a fixed background of desires, disposes the subject to behave in ways that would promote the satisfaction of his desires if its content were true*' (O'Brien 2005: 56), *across contexts*. Given that the motivational theorist places a necessary and sufficient condition on a state being a belief by appeal to its motivational role, she must also give an explanation of these features as holding only contingently (otherwise motivational role cannot be sufficient for something's being a belief). Given this understanding of belief, it looks like true beliefs are going to be more useful

to their subjects than false beliefs—perhaps the motivational theorist could appeal to this fact to explain the three features of belief outlined earlier. As Dokic and Engel, in their discussion of Ramsey, put the point:

a false belief can only be useful locally and by accident. A false belief cannot properly guide our behaviour in every possible situation. In contrast, truth systematically promotes the success of action, on a large scale. Truth offers a *guarantee* of success that falsity cannot hope to deliver. (Dokic and Engel 2002: 48)

Consider the following claims: beliefs are motivational attitudes (as per above) and true beliefs guide our actions more successfully than false beliefs (where success is cashed out in terms of desire satisfaction). Given these claims, our deliberative belief formation being characterized by Transparency, being transparent to *truth* considerations, is good news for believers. If my deliberation over *whether to believe that p* was not transparent to *whether p is true*, I might end up with more false beliefs, and thus, I may find myself engaging in less successful¹³ actions than I currently engage in. Perhaps then, we have found an explanation of Transparency—when I consider the question *whether to believe that p*, my deliberation is transparent to the question *whether p is true* because I know that having true beliefs would be more likely to bring about successful action.

However, this explanation is not faithful to the phenomenon as described by Shah, who claims that when we deliberate over whether to believe that *p*, the truth of *p* is ‘not an optional end for first-personal doxastic deliberation, providing an *instrumental* or *extrinsic* reason that an agent may take or leave at will’ (Shah 2003: 447). This explanation of Transparency implies that I have control over the considerations that my deliberative belief formation is sensitive to. If this were true then I would have to make an inference from discovering that *p* is true to determining whether I ought to believe it, perhaps via what Shah calls a ‘bridge premise’ relating to whether or not it is good to have a true belief with respect to *p*. But there is no such inferential step in doxastic deliberation, as we saw earlier, ‘there is no such gap between the two questions within the first-personal deliberative perspective; the question *whether to believe that p* seems to collapse into the question *whether p is true*’ (Shah 2003: 447). If there were such a gap, truth would be an optional end in deliberation over what to believe, but it is not, and the motivational account

¹³ This is both a qualitative and quantitative claim. That is, if my beliefs are false my actions will be less successful, but I will also engage in fewer successful actions.

does not have the resources to explain why this is the case. Belief being an attitude with a given motivational profile though *is* important to the explanation of Transparency I will offer later, it is *because* beliefs have the motivational profile that they do, that biology has made it such that deliberation over what to believe is characterised by Transparency.

Similar things might be said by the motivational theorist about why the Uncontrollability Thesis holds. Perhaps the motivational theorist could say that believing at will, without respect for epistemic considerations as in proposition (ii) of the Uncontrollability Thesis, is not going to be a very good way to collect true beliefs. Now it *might* be true that an ability to believe at will, that is, to believe something consciously, without mediation, and irrespective of epistemic considerations, is not going to be a very good way of collecting true beliefs. Even if this is right though, it does not get us an explanation of the Uncontrollability Thesis. If willed beliefs lead to unsuccessful (non-desire-satisfying) action, this does not yet tell us *why* we cannot believe at will, it only suggests that it would not be a good thing for us to be able to do. There are plenty of things that it is not a good thing that we are able to do, but we nevertheless can do them, even in the domain of cognitive states. It is perhaps not good for us that the contents of our imaginings are under our control such that we are able to imagine horrifying and frightening scenes. It is perhaps not good for us that we can desire for things to be true which are extremely unlikely to be so. We are able to do both of these things and so pointing to the fact that a given capacity would be detrimental in some way is not to offer an explanation of why we do not have it.

What of Epistemic Normativity? What might the motivational theorist say about why beliefs are governed by a norm of correctness such that beliefs with true contents are *correct* and beliefs with false contents are *incorrect*? And why is it appropriate to hold beliefs to epistemic norms? And why do we respond to such norms? These features are supposed to be distinctive of belief. How could the motivational theorist explain this?

Perhaps she could say the following: beliefs are attitudes which guide our actions. The best way for them to do this is to be true. Knowing this, we judge true beliefs as *correct* and false beliefs as *incorrect*, with the correctness here being grounded in how they fare in guiding action, and we appeal to epistemic norms in order to bring us closer to true beliefs. However, this explanation will not work because it is not by recognition of some practical norm that we judge a belief as correct or incorrect, we take the norm in play to be epistemic, and we think that it is applicable to beliefs even when they do not or cannot guide the believer's actions (see §6.5 for a similar discussion on epistemic norms applying even in cases in which a subject has no interest in forming a belief about a particular proposition). As we saw earlier, the norms supposed to be

governing belief have predominantly been understood as categorical ones. Suppose I believe that π to three decimal places is 3.147, and suppose further that I am never asked to state my belief regarding this matter, nor am I asked to carry out any calculations for which I would need to know π accurately to three decimal places. In this case, my belief is false, is it still incorrect? Clearly we think that it is. But when one judges that it is so, one's judgement is not grounded in any practical norm regarding how well this belief would guide my action, rather it is grounded in what we take to be an epistemic one. However, perhaps a belief's not guiding action is not to the point. Perhaps the motivational theorist could say that we can get the right judgement with respect to a belief's incorrectness by considering how it *would* guide action in the appropriate circumstance—so we judge it incorrect because *were* I to act on my belief, it *would* not result in successful action.

This will not work either. Now consider a subject for whom God is ensuring that actions based on a false belief are nevertheless successful in so far as their unsuccessful consequences are controlled for. So let us say that Katie has the belief that π to three decimal places is 3.147. God makes it the case that any actions Katie engages in as a result of holding this belief (and given certain appropriate desires), will have the same effects as the true belief that π to three decimal places is 3.142. So when Katie claims in a room full of π experts that π to three decimal places is 3.147, or when she writes in her maths exam that π to three decimal places is 3.147, God makes it the case that those maths experts also have a false belief on the matter, or fall asleep, or go temporarily deaf (and the same for Katie's maths teacher), so that in terms of the success of Katie's actions, it is *as if* she had stated that π to three decimal places is 3.142.

Now let us ask whether Katie's belief in this case is incorrect. It is, but this time we cannot ground this incorrectness on how the belief fares in guiding Katie's actions, because it fares perfectly well. And even if she *were* to act on the belief, it *would* guide her action perfectly well, because God is ensuring that it does. So if Katie's belief is incorrect in this case, the motivational account cannot explain this.

The motivational theorist might come back once more with something like the following: we can still ground the incorrectness of Katie's belief, even if God is looking after her, by appealing to a world in which she is not so lucky. In that world, where God is too busy, if she *were* to act on the belief that π to three decimal places is 3.147, then her actions *would* be unsuccessful. In response, let us have that it that God *necessarily* looks after Katie in this way, across worlds (this will involve huge extra theological commitments, but let us suppose that our theology was such that this kind of set up could hold necessarily). Now it is metaphysically impossible for Katie's belief that π to three decimal places is 3.147 to lead her into trouble, because it is necessarily

the case that God ensures that this does not happen. Now let us ask for a final time: is Katie's belief that π to three decimal places is 3.147 incorrect? If we think that it is—and I think that we should—then this is not something the motivational account can explain. So we have not found an explanation for Epistemic Normativity within the motivational account.

The motivational approach to belief does not have the tools to offer an explanation of the features of belief which indicate a link to truth, and there is no reason to think that it would:

[b]eing in a representational state with the right motivational powers does not obviously imply being subject in the same way to the relevant norms, and so does not obviously imply being in a state of regarding as true. (O'Brien 2005: 62)

The motivational account, though defensible, requires supplementing in order to explain Transparency, the Uncontrollability Thesis, and Epistemic Normativity.

1.5 Conclusion

In this chapter I outlined and defended a characterisation of the motivational role of belief. I then laid out what I want to explain in this thesis: Transparency, the Uncontrollability Thesis, and Epistemic Normativity. Finally, I argued that the motivational approach could not do the work required of the belief theorist in giving an account of these features.

I will take the motivational role ascribed to belief here as a necessary and sufficient condition on belief, whilst the biological function claims I make later on are contingent claims about belief (so I allow for other-world *believers*, even if they do not have biological histories, providing their attitudes play the motivational role defended here). The explanation of the three features then will be grounded on the biological functions of our mechanisms for belief-production, with the motivational part of the account in the background, used to distinguish belief from other attitudes. Unlike most philosophers in the debate, I do not take the three features which indicate a link between belief and truth to be ones which are necessary features of belief, and so they are not ones which can distinguish belief from other attitudes *across worlds*. It is the motivational role of belief, as set out here, which can do that.

2

THE TELEOLOGICAL ACCOUNT OF BELIEF

In this chapter I will consider the teleological account of belief. The teleological account takes the aim of belief to be rooted in an agent's intentions (or some sub-intentional surrogate in the case of non-deliberative belief). My own account is teleological but rooted in biology, and does not include an appeal to beliefs being *aimed* at truth in anything but a sub-intentional way.

I will argue that the explanations given by the teleological account for the three features which indicate a link between belief and truth do not work, because aims are the kinds of things which can be *weighed*, but if this is right, an explanation of the three features along teleological lines is unavailable.

The difference between the teleological account and my own comes out when we consider the nature of an aim. In the first place, if belief has an *aim* by being a (quasi) intentional activity, then one would expect believers *qua* believers to behave in a certain way, it ought to be the case that the aim of belief can be *weighed* against other aims, but this is not so with the supposed aim of belief. This problem is captured by David Owens's Exclusivity Objection, which I will be discussing. I will argue that responses to this objection do not work. These have been to identify a different place for where weighing goes on, or to replace exclusivity with something weaker: demandingness. My overall line of argument against these responses will be that, where motivational factors play a role in affecting what we believe, they do so either at the wrong point (before one enters into doxastic deliberation), or they merely change the confidence levels for the beliefs, they do *not* provide pragmatic reasons for forming or withholding belief. And these things being the case is not enough to counter Owens's Exclusivity Objection.

Another objection raised against the teleological account is that regulation by the aim is not common for all beliefs. One version of this worry is captured by the *Teleologist's Dilemma*. It has been answered by claiming that there is weak truth regulation which all beliefs share. I will argue that there is a problem with defending the claim that there is a basic level of truth regulation common to all cases of belief and that, where there is variety, there is a question mark over why these should all count as cases of belief. So the teleologist does not have the resources to classify all of these attitudes as beliefs. Indeed, other attitudes are truth regulated, guesses for example. So we need some additional reason to defend the claim that in some cases we really have cases of belief. For my account, that additional reason is provided by belief's motivational role, as defended in chapter one.

My conclusion will be that the teleological account cannot explain any of the three features which indicate a link between belief and truth, by appeal to an aim. This means that even if the teleologist is right, that belief does have an aim, this cannot do the explanatory work she takes it do. Further, objections to the teleological account have not been answered, and the account fails to capture the agent's relatively passive response to the link with truth suggestive of biological function (as I shall argue in chapter five) rather than settled by anything at which the agent can be said to be aimed.

2.1 The Truth Aim Teleological Account

The teleological account of belief holds that belief is constitutively aim governed. There are various formulations in the literature, with different claims being made with regard to what the aim of belief is. Given that I am interested in giving an account of the relationship between belief and truth, I will focus on the *truth* aim teleological account.¹⁴

The truth aim account of belief can be found in Velleman (2000)—though he has since changed his mind in this regard—and more recently has been espoused by Asbjørn Steglich-Petersen.

Before outlining the account it is worth distinguishing three notions of *aim* which I will use from here on in this chapter. Aim₁ is a conscious agent aim, the kind of aim we see in action. If I aim to run ten miles, I have an aim₁ to run ten miles. Aim₂ is a sub-conscious agent aim, this is an aim which is non-intentional, but is related to an agent's intention. It is difficult to come up with an example of this, as it is not something which we familiarly think of as a kind of aim. It is something Steglich-Petersen refers to, and we will see what this amounts to below. Aim₃ is a biological aim, one which might be derived from the function of a given trait or activity. I allow that beliefs are aimed₃ at truth, but not aimed₁ or aimed₂ at truth.

According to this account 'believing that *p essentially* involves having as an aim to believe *p truly*' (Steglich-Petersen 2009: 395, my emphasis). This aim is realised in one of two ways. In the deliberative case it is realised in the subject's aims *qua* a believer, and in the non-deliberative case it is realised by 'some sub-intentional surrogate of such intentions in the form of truth-regulated [...] mechanisms' (Steglich-Petersen 2006: 510). In the deliberate case we have an aim₁ to believe

14 At least one other aim of belief has been put forward by Conor McHugh, the aim of knowledge (McHugh 2011). Importantly though, the two objections to the teleological account I consider are concerned with the beliefs being aimed at all, not how the aim is spelled out, and so will provide traction against the knowledge aim account too. And, as Timothy Chan has pointed out: '[g]iven that knowledge entails truth, if belief aims at knowledge, it also aims at truth' (Chan 2013: 10).

p truly, and in the non-deliberative case we have an aim_2 to believe p truly. It is then, both a necessary and sufficient condition on an attitude being a belief, that it is aimed₁ or aimed₂ at truth. It is constitutive of belief that it is governed by the aim_1 or aim_2 of truth, such that an attitude's not being so aimed rules it out as being a belief. I reject the claim that belief is constitutively aim_1 or aim_2 governed (my constitutive claim relates to the motivational role of belief defended in chapter one), and I also reject the claim that belief is even contingently aim_1 or aim_2 governed, that is, I think that the thesis, even as a contingent one about beliefs in the actual world, is false. The only sense in which belief has an *aim*, is the sense in which it has an aim_3 .

2.2 Explanatory Work

In chapter one I outlined three features of belief which indicate a link between belief and truth, it is these features I seek to give an account of in this thesis. The teleological account is appealing in this respect, as it goes some way towards doing this explanatory work, though we will see later that this is not without problems. The teleological account's explanations of these features are such that they are given a strong modal reading, a reading I reject, though the modally strong versions of these theses entail the weak, and so an explanation of the former will entail an explanation of the latter. Explanations of these features as applied to beliefs across worlds will entail explanations of beliefs in the actual world.

Recall that the first feature of belief which indicates a link between belief and truth is Transparency, the fact that 'when asking oneself *whether to believe that p*, must one immediately recognize that this question is settled by, and only by, answering the question *whether p is true*' (Shah 2003: 447). Steglich-Petersen claims that if we accept the teleological account of what belief is—namely, a state which is constitutively aim_1 or aim_2 governed—we can explain Transparency with reference to the aim_1 one adopts in posing the deliberative question. If we aim_1 to believe that p with the aim_1 of accepting a truth, when we deliberate over whether to believe some proposition, Transparency 'can be explained by the aim_1 one necessarily adopts in posing that question, because the only considerations that could decide whether believing p would further that aim_1 are considerations that bear on whether p is true' (Steglich-Petersen 2008: 546). This is an explanation of the stronger reading of Transparency—because belief is *essentially* governed by the aim_1 of truth, Transparency holds necessarily.

In order to raise a worry about the kind of explanation of Transparency being offered by the teleological account, I will introduce Fred Dretske's distinction between triggering and structuring causes, which I will refer back to throughout the thesis. Dretske frames his

discussion in terms of causes of behaviour, where C is some internal event and M is the resulting bodily movement. When we are interested in the *triggering* cause of an event, we are interested in ‘what caused the C *which* caused the M’ (Dretske 1988: 42). For example, the triggering cause of my shaking hands with the Vice Chancellor is the Vice Chancellor extending his hand to me. His doing so causes the belief that the Vice Chancellor is extending his hand to me which causes me to extend my hand to him. There is, though, more to be said about the causal story here. We might want to know what the *structuring* cause of my shaking hands with the Vice Chancellor is. When we seek to find the *structuring* cause of an event we are looking for ‘what caused C *to* cause M rather than something else’ (Dretske 1988: 42). That is, we are looking for what caused me to shake the Vice Chancellor’s hand, as opposed to curtsying or waving. The *structuring* cause of my shaking the Vice Chancellor’s hand will refer to social expectation and my wishing to adhere to it. Dretske suggests that the difference between these causes gives us an account of why we might be able to know what caused C, my belief that *the Vice Chancellor is in front of me* (the Vice Chancellor being in front of me), and thus know what caused M, my shaking the Vice Chancellor’s hand (my belief that *the Vice Chancellor is in front of me*), but we can nevertheless be ignorant as to what the cause of C’s causing M is (the cause of my belief that *the Vice Chancellor is in front of me* causing my shaking his hand (rather than something else)) (Dretske 1988: 43).

Very briefly: my preferred account gives an explanation of Transparency which appeals to both the triggering and structuring causes. The rough idea (to be discussed fully in §5.3.1) is that the triggering cause of Transparency can be given by appeal to various kinds of brain development, so there are neurological structures which realise the disposition to allow certain considerations to settle whether or not one should have a certain belief. The structuring cause can be given by appeal to natural selection, which selected for those neurological structures. ***

Returning to the teleological account’s explanation of Transparency, this is given by appeal to the aim₁ one adopts when one enters into deliberation about what to believe. So perhaps this is an explanation which cites the aim₁ as a triggering cause. What explains the disposition to allow certain considerations to settle whether or not one should have a belief that *p*, to move from the question *whether to believe that p*, to the question *whether p is true*, is the aim₁ one adopts in posing the deliberative question. Note that when we locate the role of truth at an intentional level—with the subject aiming₁ to believe a truth—we only get an explanation which gives us a triggering cause of Transparency, we get an explanation of each case, not an explanation which explains each case, not an explanation of what structured the disposition to move from certain considerations to belief, to move from the question *whether to believe that p*, to the question *whether p is true*. The teleological account is silent as to what caused the aim₁ which causes Transparency;

the aim₁ which is supposed to be explanatory is the triggering cause of Transparency, no structuring cause is offered (why are we so aimed₁?). Perhaps the teleological account is compatible with a biological approach, the aim₁ is the triggering cause of the disposition to be moved from to move from the question *whether to believe that p*, to the question *whether p is true*, and indeed from *whether p is true* to *belief that p*. What secures why we are so aimed₁ is biology. However, if the aim₁ is the triggering cause, presumably we can give a lower level causal description of what is going on, a causal description which includes reference to the neurological structures which secure Transparency, and then talk of aims₁ becomes explanatorily redundant. On the other hand, if the neurological structures which secure Transparency are accepted as the triggering cause, then we only need to appeal to biology as the structuring cause, there is no explanatory role for the aim₁ to play. Further, even if we *wanted* to make a place for an aim₁ in the causal picture, the aim *cannot* come into the account as the structuring cause, because this seems to get things in the wrong order. It does not look plausible to claim that the aim₁ of belief is the structuring cause of Transparency, that the aim₁ *causes* the neurological structures *to* cause the disposition to move from the question *whether to believe that p*, to the question *whether p is true* or from *whether p is true* to *belief that p*. How could a subject's aim shape neurological structures in this way? When we are thinking in terms of triggering and structuring causes then, the notion of an aim₁ does not sit comfortably in this story.

Another more general problem with appealing to aims₁ is that they are not at the required strength to do the explanatory work, because aims₁ are the kinds of things which can be weighed against one another (I will defend this claim in §2.3). Belief could be governed by the aim₁ of truth without it being the case that Transparency holds. It is consistent with belief have an aim₁ that other considerations could weigh into the deliberation which results in the fixation of belief. Transparency is the immediate and non-inferential collapsing of one question into another, and so it cannot be an aim₁ which explains this because aims₁ allow for weighing. If it were an aim₁ of mine which explained why I moved from the question *whether to believe that p* to *whether p is true*, I ought to be able to resist this move. But Transparency does not allow me to do this. The teleological account then, for two reasons, cannot explain Transparency.

Let us move on to the teleological account's explanation of the Uncontrollability Thesis. The explanation is of a strong version of the Uncontrollability Thesis which holds as a matter of metaphysical necessity. If one were to form a belief at will, one would be doing so without regard for whether it were true or not, and would not therefore be forming it with the aim₁ of accepting a truth. Believing at will would be to form a belief 'without the aim₁ requisite to its being a belief' (Velleman 2000: 244). So again, because belief is *essentially* governed by the aim₁

or aim₂ of truth, if one forms a belief irrespective of considerations of its truth, one would not be forming a *belief*, because the resulting state would not be governed by the aim₁ necessary to belief—the aim to believe that *p* only if *p* is true.

However, note again that appeal to an aim₁ cannot do the explanatory work here. Belief could be constitutively governed by the aim₁ of truth, without it being the case that it is impossible (necessarily or contingently) to believe at will. Having one's believing governed by a particular aim₁ does not mean that one cannot believe something which is not in line with that aim₁—it is consistent with something's being governed by a certain aim₁ that other considerations could weigh into the deliberation which results in the fixation of belief.

Turning now to Epistemic Normativity, I said in the last chapter I would understand this as constituted by two claims: (EN1) Beliefs have a standard of correctness, and (EN2) Beliefs are governed by categorical epistemic norms. I suggested that the truth of (EN1) and (EN2) requires an explanation, or at the very least, what needs an explanation is why we think that (EN1) and (EN2) are true. With respect to (EN1), the proponent of the teleological account has an explanation: 'believing *p* is correct only if *p* is true because only true beliefs achieve the aim₁ involved with believing' (Steglich-Petersen 2009: 395). Note here that whereas I formulated the correctness condition earlier in terms of sufficiency, here it is formulated as an *only if*. The explanation given for belief's standard of correctness here applies just as well to this standard as a sufficiency claim. A belief that *p* is only correct when *p* is true (we get this from Steglich-Petersen's explanation), but also, when *p* is true, a belief that *p* is correct. Both sides of the conditional look equally explainable by appeal to the claim that true beliefs achieve the aim₁ involved with believing. There is no worry of demanding too much of believers with having the correctness condition in terms of sufficiency, because it is merely descriptive (compare with the Too Demanding Objection to the normative account, §3.2). On the teleological account, *p* being true does not indicate that you *ought* to believe it, but only that, if you are going to form a belief on the matter, it would be correct to believe that *p*, and if it is correct to believe that *p*, *p* is true. On the teleological account, no prescription is derived from belief's standard of correctness (one is not *obliged* to believe *p* only if *p* is true, for example, it is only that it is correct to believe that *p* if and only if *p* is true, in line with our aim₁ as believers).

With respect to (EN2), the teleologist can explain the existence of epistemic norms by claiming that 'following them promotes the aim₁ of believing truly' (Steglich-Petersen 2009: 396). If aims₁ have rules or standards associated with the achieving of them, then epistemic norms might be considered the rules or standards conducive to achieving belief's aim₁ (McHugh 2011: 371). Epistemic norms, on this account, are instrumental ones, 'deriving their normative

force from their ability to guide us to achieve our aims_[1]' (Steglich-Petersen 2009: 396). The teleologist then will reject my characterisation in (EN2) of epistemic norms as categorical. Epistemic Normativity has its source in the desire of the believer to believe truly, there 'seems to be nothing more to its being good when I believe the truth, and bad when I don't, than the fact that believing the truth was what I was trying to do' (Hazlett 2013: 162).

The nature of belief, on the aim account, is to be understood only descriptively, there is no need to appeal to irreducible epistemic norms (Chan 2013: 3). The aim theory, it is claimed, promises 'a simple, unified, and *prima facie* unproblematic explanation of epistemic normativity' (Steglich-Petersen 2009: 396).

However, the following worry arises at this point: from (EN2) we learn that beliefs are governed by categorical epistemic norms, or at the very least, we think that they are. But according to the teleologist, such norms are instrumental ones, which derive their force from their being able to guide us in achieving the aim₁ of belief. But we *take* these norms to be categorical, we take them to apply to us regardless of our interests (see §6.5). If they were simply instrumental norms indicating how best to achieve our aim₁, we ought to (at least be *able* to) regard them as instrumental. The reason we cannot do this is because the aim₁ of belief does not lend itself to weighing, as we will see in the next section. This is problematic because it is part of the nature of aimedness₁, that an aim₁ can be weighed against or constrained by other aims₁ a subject might have. If this were possible, the claim that epistemic norms were instrumental ones serving to indicate how best to achieve the aim₁ of belief, would look more plausible.

We have seen then that the teleological account's explanation of Transparency is problematic insofar as, when we think in terms of triggering and structuring causes, the aim₁ which is claimed to be explanatory, does not comfortably sit in this picture. Also, a more general point about appealing to an aim₁ is that this cannot explain any of the features, because we cannot weigh the aim₁ of belief (as I will argue in §2.3). We cannot explain Transparency by appeal to the aim₁ we have in forming a belief, because Transparency is the immediate and non-inferential collapsing of one question into another, this collapsing is not something we can resist by weighing in other aims (as we ought to be able to do were it an aim₁ which explained Transparency). As for the Uncontrollability Thesis, we cannot explain why we cannot believe at will by claiming that doing so would not be conducive to achieving the aim₁ we have in believing, because once again, were we able to *weigh* the aim₁ against others we have, then there should be cases in which we forego the aim₁ of belief, and believe at will in line with some other aim₁ we have. We have also seen that the explanation of Epistemic Normativity is problematic insofar as it does not respect our taking epistemic norms to be categorical—if they were merely

instrumental devices we can call upon to help us achieve our aim₁, we ought to be able to weigh them. But we cannot. My argument here relies on the aim₁ of belief not being weighable against other aims₁, next I will defend this claim with reference to Owens's Exclusivity Objection.

2.3. The Exclusivity Objection

Owens put forward the Exclusivity Objection to the truth aim account, in which he claims that the aim₁ of belief, as offered, does not cohere with our ordinary notion of aimedness₁. We should understand Owens's argument as directed against the claim that beliefs are aimed₁ (and derivatively aimed₂ if being aimed₂ depends on some aim₁), it does not speak against beliefs being aimed₃. Talk of beliefs having aims₁ then, becomes in danger of being metaphorical. Importantly, Owens attempts to elucidate exactly what it might mean for belief to have an aim₁: he offers a necessary condition to be met if they are so aimed₁, and argues that they do not meet it.

Owens formulates the truth aim₁ hypothesis thus: 'Ø-ing that *p* aims₁[.] at the truth if and only if someone who Øs that *p* does so with the purpose of Ø-ing that *p* only if *p* is true' (Owens 2003: 289). Owens notes that he has formulated the truth aim₁ hypothesis negatively; that is, an agent is only required by it to avoid error, she need not seek to believe all or any particular truths (Owens 2003: 289). This is important: if the claim of the truth aim₁ account was that we, *qua* believers, aim₁ to believe that *p* if and only if *p* is true, we would find ourselves attempting to meet an impossible aim₁ (there are simply too many truths—some of which are inaccessible to human believers¹⁵). I will argue later that given this very point, Steglich-Petersen's examples given in response to this objection, do not work.

Owens claims that the explanatory value of the truth aim₁ hypothesis lies in its explaining of the correctness condition of beliefs (which I identified as claim (EN1) of Epistemic Normativity), as well as the role that evidence plays in justifying ones beliefs (Owens 2003: 295). However, Owens argues that the aim₁ of believers as given by the teleologist does not cohere with our ordinary notion of aims₁, that is, we understand aims₁ as things capable of interacting with and being weighable against each other, this feature of aims₁ is necessary to them. My aims₁ can be weighed against, and constrained by, one another. It is essential to aims₁ that they are 'the kind of things that can conflict and be weighed against each other' (Steglich-Petersen 2009: 396), such that if a state cannot be weighed in this way, then

15 As analogue of this kind of worry arises for the formulation of the norm supposed to be governing belief by the normative account (§3.2).

it is difficult to seriously think of that state or consideration as an aim_[1]. If an act is not the outcome of aims_[1] being weighed against each other, it is difficult to think of that act as aim_[1]-directed. (Steglich-Petersen 2009: 396)

For example, my aim₁ to stay up late working is constrained by my aim₁ to rise early for a meeting. What Owens is up to here is placing a constraint upon aims₁ which is derived from the aims₁ we have in acting. In this domain, the aim₁ one might have with respect to some activity can be weighed against one's other aims₁. It is worth noting at this point that my preferred account—even though it is teleological—does not face this kind of worry. This is because I explain the link to truth by appeal to one of the biological functions of the mechanisms which produce beliefs, not by appeal to any aim₁ of an agent. This means that there is no problem for my account posed by the exclusivity of truth considerations in deliberation over what to believe, because I do not explain this exclusivity by appeal to an aim₁, but rather, by appeal to an aim₃, secured by biology.

So if we want to give an explanation of our adherence to epistemic norms for example, and if we want to do this by appeal to some aim₁ we have in believing, 'we should expect that we, on at least some occasions, when deciding how to adhere to the norms generated by that aim_[1], weigh the aim_[1] of belief against other aims_[1]' (Steglich-Petersen 2009: 396).

Owens's claim is that this necessary condition on aimedness₁ is not met by the proposed aim₁ of belief; the aim₁ of truth. If it really were the case that I aimed₁ to believe *p* only if *p* were true, then, given the nature of aims₁, I could weigh my aim₁ to believe truly, against say, my aim₁ to be happy. One might weigh one's aim₁ to have a true belief regarding one's spouse's fidelity against one's aim₁ to keep peace within the relationship, and, if the latter aim₁ is accorded more weight, one may end up falsely believing that one's spouse is faithful. However, as we have seen, our deliberation over what to believe is characterised by Transparency, which rules out this kind of weighing. When we deliberate as to *whether to believe that p*, we find that we can only be moved by evidential considerations as to *whether p is true*, regardless of any non-epistemic aims₁ we might have when we enter into this process. Owens claims that this exclusion of our other aims₁ which occurs in deliberative belief formation, shows that belief is not an aim₁ governed activity; that is, there is no aim₁¹⁶ of belief. If there is no aim₁ of belief, we cannot appeal to one in an

16 We might only get something weaker from Owens's argument: it might be that there is no *truth* aim of belief. There might be some other aim which is weighable. However, it is enough for my purposes to show that there is no aim in the way teleological theorists have supposed, that is, there is no aim which can be appealed to in order to explain the connection between belief and truth.

explanation of the three features of belief which indicate a link between it and truth, and so the teleological account does not represent an alternative to my preferred view.

2.3.1 Responses to the Exclusivity Objection

McHugh claims that there are three ways the teleologist might go in responding to the Exclusivity Objection. One way would be to give an explanation of exclusivity which is compatible with the claim that beliefs have an aim₁. The second is to deny that the teleologist is required to explain exclusivity, and the third is to deny exclusivity altogether, and to give an account of the phenomena which look to indicate exclusivity in terms compatible with the teleological account (McHugh 2012: 436). With respect to the second kind of response—that the teleologist is not required to explain exclusivity—I am going to put this aside. In this chapter I am assessing the teleological account with respect to its ability to give an explanation of the three features of belief outlined in the previous chapter. If the teleologist claims that she need not give an explanation of exclusivity to truth considerations in doxastic deliberation, then she is not giving an account of the connection between belief and truth, and so she is not giving an account which needs to be discussed here. In this section then, I will discuss the first and third kinds of response, and argue that they do not work.

2.3.1.1 Accepting Exclusivity

The teleologist could respond to the Exclusivity Objection by accepting that deliberation over what to believe is governed exclusively by truth considerations, but that this is compatible with belief having an aim₁. We find this strategy in Steglich-Petersen's response to the objection.

Steglich-Petersen argues that the truth aim₁ for belief meets Owens's necessary condition on aimedness₁, insofar as it interacts with other aims₁ in the appropriate way. He accepts both Owens's condition on aimedness₁, and importantly, Owens's construal of the truth aim₁.

A small digression is required here with regard to Steglich-Petersen's acceptance of Owens's construal of the truth aim₁. It is worth noting that Steglich-Petersen, at one point, states the truth aim₁ using sufficiency, claiming that '[i]n forming a belief in a particular proposition, one aims₁ at accepting the proposition in question if and only if that proposition is true' (Steglich-Petersen 2009: 402). However, Steglich-Petersen does not claim that Owens's characterisation of the truth aim₁ is incorrect, and as such, I will consider this sufficiency claim as either a slip, or as referring to what an agent is doing once she enters into deliberation about whether or not to believe that *p*. Formulating the truth aim₁ as Owens does will not just be a necessary feature of Steglich-Petersen's approach, but all approaches of belief which appeal to a

truth aim₁. A truth aim₁ characterized in terms of sufficiency is unacceptable because it attributes to believers aims₁ they do not have (see §3.2 for an analogue of this worry for the normative account of belief).

Steglich-Petersen claims that Owens failed to locate evidence that the truth aim₁ is weighed, because he only considered one way of its being so. Owens only considered the weighing of aims₁ in which one '[combines] them in action', and failed to consider cases in which one weighs one's aims₁ by discarding one of the aims₁ in question (Steglich-Petersen 2009: 401). Of course, he says, if an 'aim₁' fails to meet even this standard, then we should stand with Owens in claiming that due to this failure, it can not be thought of as a genuine aim₁ (Steglich-Petersen 2009: 401).

However, Steglich-Petersen argues, the truth aim₁ for belief can be weighed in this second way—the result of which is that a contrary aim₁ is discarded as a result of mutual incompatibility. For example, an agent may weigh the aim₁ of belief against her other aims₁ in cases where she considers whether or not to even take up the aim₁ of belief, that is, whether or not to even begin the process of forming a belief about *p*. The agent weighs (the carrying out of) the aim₁ of belief against any non-epistemic aims₁ that she has. Equally, there are certainly cases where the agent gives up the aim₁ of belief in regard to *p* (by not forming a belief about *p*) because this aim₁ is in conflict with her other aims₁ (Steglich-Petersen 2009: 402–3). To take an example; I may consider it to be a poor use of my time and/or my cognitive resources to find out whether or not there are more than *n* blades of grass in my garden. Here then, I weigh the aim₁ of belief against my other aims₁ (using my time wisely, saving my mental energy for important tasks), and decide not to form a belief about how many blades of grass are in my garden, and thus discard the aim₁ of belief.

Steglich-Petersen also claims that one can give up the aim₁ of belief with regard to a given proposition, even when the aim₁ has already been adopted. Cases in which agents decide to suspend belief about *p* given inconclusive evidence or the apparent incompatibility with their other aims₁ that forming a belief about *p* may cause, exemplify the giving up of the already adopted aim₁ of belief (Steglich-Petersen 2009: 403). We can see this latter case with an example: a classroom window is broken and so a teacher may decide to find out which student was responsible. However, she realises that in forming such a belief she would have to punish the guilty party, the prospect of which she finds unpleasant. As such she decides not to form a belief about the responsible child and thus gives up the aim₁ of belief (Steglich-Petersen 2009: 403). The claim is that here we have a case of non-epistemic, practical aims₁ being weighed against the aim₁ of belief and the former winning out, with the latter being discarded.

Steglich-Petersen's response works like this: he accepts that for something to be an aim₁ it is necessary that it is able to be weighed against other aims₁. But, there are two ways in which this weighing can be done. The first is by combining the aims₁ resulting in action; this is not what happens with the aim₁ of belief. However, we can see that the other way of weighing aims₁ does occur in the case of the aim₁ of belief. The aim₁ of belief can be weighed against other aims₁ and discarded in favour of pursuing those other aims₁ (or vice versa). As such Owens is wrong when he claims that the aim₁ of belief does not allow for the requisite weighing shared by all others aims₁, and so too falls down his contention that the truth aim hypothesis fails.

There are two problems with Steglich-Petersen's response to Owens's objection.¹⁷ The first has to do with the examples he uses to demonstrate that the truth aim₁ can be weighed against other aims₁, these are problematic because they employ a conception of the truth aim₁ which is not shared by Owens—and, interestingly—claimed to be avoided by Steglich-Petersen himself. Recall that Owens gave a negative formulation of the truth aim₁ which meant that it did not require agents to believe all truths, rather, it only required the avoiding of falsehoods. Given this, it is not the case that the truth aim₁ requires of a subject that she comes to hold a belief about the number of blades of grass in her garden, nor does it require of the school teacher that she come to hold a belief about who broke the classroom window. Therefore, the right thing to say in these cases is not that the agents weigh the truth aim₁ and discard it in favour of practical considerations, rather, the truth aim₁ does not require consideration. This is because, to repeat, agents are not required by the truth aim₁ to form beliefs, rather, it is only that, *if* that is what an agent is up to, *then* the beliefs which she comes to have had better be true ones.

The examples Steglich-Petersen offers for the claim that the truth aim₁ can be weighed look to be taking the truth aim₁ as a sufficiency claim: that is, the truth aim₁ is construed as proposing that beliefs are aimed₁ at the truth only if they should be formed in response to any truth. I noted earlier that Steglich-Petersen makes this slip. Now, if the truth aim₁ was to be understood as making a sufficiency claim, then, of course, this aim₁ could be weighed against other aims₁ an agent may have. But Owens noted—as did Steglich-Petersen—that this is not the way the truth aim₁ is to be understood.

The second problem with Steglich-Petersen's response has to do with his equivocating between deliberating over whether to *form a belief about p*, and deliberating over whether to *believe that p*. The former is not a part of the activity of forming a belief that *p*. If a subject is

¹⁷ The two problems I raise in the rest of this section appear in a paper written with Paul Noordhof (see Sullivan-Bissett and Noordhof 2013).

deliberating about whether to form a belief about p (*should I form a belief about who broke the classroom window?*), then she is not yet in the business of forming a belief about p (*who broke the classroom window?*). There is no adoption of the truth aim₁ when an agent enters into deliberation about whether to form a belief about a given proposition. An agent can arrive at an evaluative proposition regarding whether or not to form a belief about p , but in doing so, she weighs the aim₁ of truth *without adopting it*, she deliberates over whether or not to *take up* that aim₁. If an agent decides to form a belief about p (or not), she arrives at this on the basis of the evaluations she comes to which arise by deliberating about whether to form a belief about p . When an agent in these circumstances decides not to form a belief about p , Steglich-Petersen claims that she discards the aim₁ of belief. But this is the wrong way to describe the case. What is going on is that the decision not to form a belief about p has been informed by a cognitive process prior to that of belief formation. She has not already adopted the truth aim₁ for p . So it is not weighing one aim₁ against another. In deciding whether to form a belief it is not even that the value of truth is being weighed against other aims₁. Rather, truth is a constraint upon what proposition one believes if one is in the business of forming a belief about a subject matter.

I am not denying that one can weigh the truth aim₁, but this in itself is not enough to respond to Owens's objection. Such weighing has to take place in the right context; the context of belief formation. When one thinks about whether to engage in the process of belief forming, one is in a different state and this is an inappropriate context for weighing the truth aim₁ because the aim₁ must be the aim₁ of *belief* not *belief forming*. The weighing process when one weighs the aim₁ of truth is with respect to whether the agent should form a belief *about p*. Steglich-Petersen needs the weighing to be concerned with whether to believe *that p*, if he is to answer Owens's objection.

Steglich-Petersen's response to Owens's objection fails for two reasons, and so the teleologist needs to look elsewhere for a response to the Exclusivity Objection.

2.3.1.2 Denying Exclusivity

We have seen one response to the Exclusivity Objection which accepts that deliberation over what to believe is characterized by exclusivity, and claims that the aim₁ of belief can nevertheless be weighed against an agent's other aims₁, this was a version of what McHugh identified as the first possible strategy (McHugh 2012: 436). I argued that this response does not work. We can now look to the possibility of denying exclusivity as a way of meeting Owens's objection, this is what McHugh identified as the third possible strategy (McHugh 2012: 436).

McHugh opts for the third strategy and claims that the ‘data that generate the impression of exclusivity can be explained in another way’, and that this way is compatible with the claim that the aim₁ of belief can be weighed (McHugh 2012: 436). Instead of exclusivity, we should recognise a feature of the regulation of belief which McHugh calls *demandingness*, which captures the fact that

you cannot deliberately form a belief in a proposition if you regard your evidence for that proposition as less than *sufficient*, where sufficiency involves more than having better or stronger evidence for the proposition than for its negation. You require what you take to be some high degree or strength of evidence, or some particular kind of evidence, for the proposition. (McHugh 2012: 436)

Exclusivity and demandingness are different phenomena, the latter allows for the kind of weighing which would make coherent the claim that belief is aim₁ governed. McHugh’s strategy is to deny exclusivity and instead account for the cases in which exclusivity is supposed to play a role, by appeal to the feature of demandingness. If deliberative belief formation is governed by demandingness, we can claim that belief has an aim₁. The difference between exclusivity and demandingness is that the former rules out non-evidential considerations being reasons for belief within deliberation, but says nothing about the strength of evidence required to form a belief. The latter, on the other hand, allows that non-evidential consideration can be reasons for belief, but it requires that ‘if you regard your evidence as insufficient, such considerations cannot effectively motivate you to go ahead and believe’ (McHugh 2012: 437). So if non-evidential considerations have motivational force, if they are derived from some aim₁, this is one other than the aim₁ of truth (McHugh 2012: 437). Demandingness allows for the aim₁ of belief to interact with other aims₁ when a subject is deliberating over what to believe, and if it is demandingness which characterizes doxastic deliberation—and not exclusivity—then Owens’s challenge is met. I will flag at this point that I think that it is odd to call these non-epistemic considerations *considerations* at all, given that no amount of them can outweigh or make up for evidential shortfalls.

If demandingness characterises our belief formation, this has implications for the explanatory project I outlined in chapter one. If McHugh is right, then it is not the case that our deliberation over *whether to believe that p* is answered only by answering the question *whether p is*

true, that is, Transparency does not characterise our deliberative belief formation.¹⁸ Instead what needs an explanation is why our deliberative belief formation is characterised by demandingness, and this explanatory work can be done by appeal to belief having an aim₁. If belief is constitutively aim₁ governed, we can explain demandingness by reference to the aim₁ one adopts when entering doxastic deliberation. One cannot form a belief that *p* deliberately if one regards the evidence for *p* as less than sufficient, because forming a belief that *p* on less than sufficient evidence is not conducive to the aim₁ one has as a believer.

There are cases in which exclusivity and demandingness make the same prediction. In such cases deliberation being characterised by demandingness is just as good an explanation of our belief-forming practices as its being characterised by exclusivity, but according to McHugh, there are cases in which demandingness is in play, while exclusivity is not. Cases in which demandingness and exclusivity predict the same thing are those in which a subject considers whether or not to believe that *p*, and she regards as insufficient both her evidence for *p* and her evidence for $\sim p$. In such a case the subject ‘will in effect be forced to withhold belief, no matter how glorious the riches [she is] offered to believe *p* (or $\sim p$), or what other non-evidential considerations are in play’ (McHugh 2012: 438). This is a prediction licensed by both exclusivity and demandingness.

What McHugh needs for his claim that doxastic deliberation is characterised by demandingness and not exclusivity are cases in which an appeal to demandingness explains the case, without appeal to exclusivity, or indeed, cases in which it looks like exclusivity is not in play. McHugh offers several cases, I will consider two here. The first case is one in which a friend stands accused of some terrible wrongdoing, and his having committed the wrongdoing has been reported by several witnesses. For you to have this evidence, and your friend be innocent, it would have to be the case that all of the witnesses were lying or had made a mistake, a distant possibility, but one that cannot be ruled out, given your evidence. Further, you value the friendship you have with the accused, and recognise that believing him guilty would damage the friendship (especially if he were innocent). In this case, McHugh says that though

you cannot clear-headedly believe that your friend is *not* guilty [...] it seems that, in deliberating about whether to believe that your friend is guilty, you can take into account a non-evidential consideration, namely the potential damage of believing that your friend is guilty, and you can

¹⁸ McHugh thinks that if demandingness characterizes our belief formation, as opposed to exclusivity, then the phenomenon of Transparency identified by Shah does not characterize belief formation, and the motivation for the normative account of belief is lost: ‘if normativism entails [T]ransparency, then the falsity of transparency would show that normativism is false’ (McHugh forthcoming: 16).

conclude deliberation by withholding belief, remaining open to the possibility that the witnesses are mistaken or lying, for this reason. (McHugh forthcoming: 9)

Here is a case then in which a subject can withhold belief for a non-evidential reason. If deliberative belief formation were characterized by exclusivity, in this case the subject could not withhold belief in her friend's guilt on the grounds that believing her friend guilty would be damaging to the friendship, it is only if the evidence were equal (by the subject's lights) that she could withhold belief. Demandingness though allows this. The proponent of exclusivity needs to say that this way of characterising the case is a mistake.

The second case McHugh offers to show that exclusivity does not characterise doxastic deliberation is the following: you have a cheque which needs to be cashed by the end of the week. You know it would be easier to cash it on Saturday than today, which is Friday. You also know that the bank has been open on Saturdays before, and so you have sufficient inductive grounds to believe that the bank will be open this Saturday, even though you cannot rule out the possibility that the bank has changed its opening hours, and so might be closed on Saturday. Though you would usually form the belief that the bank is open, given that the stakes are high in this case, it 'seems that you can, for this reason, withhold belief on the matter, taking seriously the possibility of a change in hours' (McHugh forthcoming: 9).

Again, we have a case in which a subject can withhold a belief for a non-evidential reason. And again the exclusivity proponent has to say that this kind of case is impossible. So it looks like it is not exclusivity which governs belief formation, but rather demandingness. And if that is right, the aim₁ of belief can be weighed in certain circumstances, and Owens's exclusivity objection is answered.

The first thing to note about McHugh's response is that withholding belief is not a case of weighing pragmatic and epistemic aims₁. We can see this by looking to the way the aim₁ of truth behaves when it governs guessing, we do not get the kind of result we get in the belief case. Guessing are governed by the aim₁ of truth; the standard of correctness for a guess is truth, and someone cannot be described as guessing if they are not intending to guess truly (Owens 2003: 291). However, it is not the case that epistemic considerations always trump deliberation, such that the best we can do is withhold a guess. Consider Owens's case:

[s]uppose the quiz master asks me whether the earth's population is greater than 7 billion. There is \$1 million at stake. I wouldn't claim to know the answer but no answer means no money. Obviously I ought to make a guess. So I plump for the answer 'yes'. My guess is a sensible one, even though I have no evidence on the matter, because if I don't guess at all I am certain to lose

the money. In fact, a guess would be sensible even if I had to get the population of the world to within ten million. Here I'm pretty unlikely to be right but, so far as my evidence goes, any guess I make is as good as any other and I have to make some guess to win. (Owens 2003: 291–2)

Here we have a case in which though truth is the aim₁ of guessing, non-evidential considerations interact with the evidential ones, this of course, *never* happens with belief. All we get in the belief case is our being able to *withhold* belief in light of pragmatic considerations. McHugh's cases are important though, they raise the question of what is going on in these cases, where pragmatic considerations do look to be affecting the fixation of belief. Now, it is not the case that there is a weighing of such considerations going on, and so these kinds of cases do not help in a response to Owens, but they do demand an explanation.

We do not need to concede that practical considerations can influence belief formation in the way described, but rather we can understand these considerations as functioning to modify the standards for sufficient evidence required for belief, and not as reasons for the subject to withhold belief. We might say something like the following: what these practical considerations do is 'help to fix *how* easily a proposition's being false despite your evidence counts as *easily*, such that the evidence is not sufficient' (McHugh forthcoming: 10). So, when the stakes are high, as in the cases McHugh describes, your standards for sufficient evidence are high, and so your being moved to belief requires more or better evidence. So the reasons which motivate withholding belief are purely evidential, in line with exclusivity.

McHugh canvasses this line of response. In reply he claims that even if practical considerations can affect what counts as sufficient evidence for a subject, this does not show that these considerations cannot motivate a subject to withhold belief. In fact, what might make a practical consideration a good candidate for a reason one might withhold belief is its playing a role in raising the subject's standard for sufficient evidence (McHugh forthcoming: 10). In order for the response to McHugh's cases to be problematic for the way in which he describes them, it would have to be the case that though a practical consideration can raise a subject's standard for sufficient evidence in some context, it cannot be a *reason* for which the subject withholds belief. So though a practical consideration can influence a subject's withholding a belief, it does so *without being a reason for which the subject does so*. McHugh suggests that it would be 'puzzling what would be going on' from the subject's point of view in such a case, given that one 'cannot deliberately acknowledge the role of the practical consideration in such events of doxastic rearrangement, and yet there is no change in [the subject's] evidence that could explain them either' (McHugh forthcoming: 10). So the claim is that in order for the response to work, it would need to be the case that practical considerations can affect a subject's standard for

sufficient evidence, without those considerations being a *reason* for which the subject withholds belief. My claim is that this is exactly right. Non-epistemic considerations on confidence levels for belief are not recognised by the subject *as reasons* for withholding belief in certain contexts, the effect they have is not recognised in deliberation over what to believe.

Once we distinguish between non-epistemic reasons for withholding belief, and the influence of non-epistemic interests on confidence levels for belief, we see that the putative cases of demandingness are really cases of exclusivity, and further support the claim that exclusivity characterises our deliberative belief formation. So the way we should understand McHugh's cases is not as cases in which non-epistemic considerations function as reasons for withholding belief, but rather as cases in which non-epistemic considerations change what the subject requires as sufficient evidence to form a belief.

We can look to Yaacov Trope and Akiva Liberman's (1996) idea of confidence thresholds for belief. The idea here is that the lower a subject's confidence threshold, the less evidence is required to reach the threshold. The *acceptance threshold* is defined as 'the minimum confidence in the truth of a hypothesis that [one] requires before accepting it, rather than continuing to test it' whilst the *rejection threshold* is 'the minimum confidence in the untruth of a hypothesis that [one] requires before rejecting it and discontinuing the test' (Trope and Liberman 1996: 253, cited in Mele 2000: 34). These thresholds depend on 'the cost of false acceptance relative to the cost of information' and 'the cost of false rejection relative to the cost of information' respectively (Trope and Liberman 1996: 253, cited in Mele 2000: 34). By *cost of information* what is meant is the resources and effort that is required for the subject to acquire and process information relevant to the target proposition. By *cost of false acceptance* and *cost of false rejection* what is meant is the subjective importance the subject attaches to avoid falsely believing a proposition and avoid falsely believing the negation of a proposition, respectively (Trope and Liberman 1996: 252, cited in Mele 2000: 34). If this model is right, our desires can influence our beliefs, though not in the way that McHugh suggests. Rather, they function to change the thresholds, but are not recognized as *reasons* to withhold belief.

We can also look to Jason Stanley's Interest Relative Invariantism about knowledge, to support the claim that it is changes in confidence levels which are affected by non-epistemic considerations which account for McHugh's cases, as opposed to non-epistemic considerations directly affecting belief formation, functioning as reasons to withhold belief. According to Interest Relative Invariantism

knowing a true proposition requires a subject who believes it to possess a sufficient level of evidence for that proposition, where sufficiency is measured in terms of some kind of probability [...] the greater the practical investment one has in a belief, the stronger one's evidence must be in order to know it. (Stanley 2005: 88)

Stanley also discusses a version of the bank case. When the subject does not have a cheque to cash, she is in a 'Low Stakes' situation, and so the proposition that the bank is open on Saturday does not pose a serious practical question for her. In a 'High Stakes' situation, when the subject has a cheque to cash, the proposition is a serious practical question for her, given that if the proposition is false (if the bank is not open), the warranted expected utilities of some of the actions at the subject's disposal are affected (namely, the utility of going to the bank on Saturday) (Stanley 2005: 96–7). For Stanley, in a High Stakes case, the subject does not have enough evidence to know the target proposition.

We can understand McHugh's cases in a similar way. We might think that the standards required to be a believer are modified according to context. In High Stakes cases, as in the cases McHugh describes, the warranted expected utilities of the subject's actions are affected if they falsely believe the target proposition. These are not cases in which non-epistemic considerations directly influence belief formation, but are rather cases in which the subject requires more evidence to be a believer. This does not speak against doxastic deliberation being characterized by exclusivity, but rather vindicates this claim. When non-evidential considerations affect deliberative belief formation, they only do so by changing the standards of evidence required, not by functioning as a consideration or reason for withholding belief.

To conclude this part: I argued that withholding belief is not to weigh the aim_1 of belief. I then argued that it is not demandingness that characterises deliberative formation, and that McHugh's cases are better characterised as cases in which non-epistemic factors influence the standards of evidence required, but do not enter doxastic deliberation as considerations which are weighed in the fixation of belief or withholding belief. Owens's Exclusivity objection shows that belief does not have an aim_1 and remains unanswered.

2.4 The Teleologist's Dilemma

I will now move on to discuss a second problem with the teleological account, the fact that regulation supposedly secured by the aim_1 and aim_2 of belief is not common for all beliefs. This objection is expressed by Shah's *Teleologist's Dilemma*. We have seen that according to the teleological account belief is governed by an aim, realised in the intentional aims₁ of the subject,

or some sub-intentional surrogate of the aim₁ (an aim₂). Shah argues that the teleologist, in claiming that an intention to accept a truth is a necessary condition for a state to count as a belief, finds herself on the horns of dilemma (Shah 2003: 461). This is because the teleological account needs to say something about non-deliberative beliefs; that is, beliefs which are not formed as a result of our aiming to accept a truth. So deliberative beliefs are subject to Transparency, that is, when one considers *whether to believe that p*, it is only considerations of truth that play a role in answering that question. But there are beliefs which do not come about through deliberation and are not subject to Transparency. We have seen from Steglich-Petersen (and Velleman), on the teleological account there are some mechanisms, or a disposition which constitutes aiming₂ at the truth in the absence of the believer's conscious intention. Shah claims that the teleologist must give an account of this disposition but she must do this in a way which allows non-deliberative beliefs through the net. This may involve casting such a disposition as rather weak, as in some cases of non-deliberative beliefs, the role of truth played looks like a weak one. However, the teleologist must also account for why, when one forms a belief through the process of doxastic deliberation (which is after all, just *one way* in which one can form a belief¹⁹), truth is the only relevant consideration for the agent; to the exclusion of all other considerations. She could do this by construing the disposition which constitutes aiming₁ at the truth as particularly strong, but then she would find herself unable to account for weakly truth-regulated non-deliberative beliefs (Shah 2003: 461–3).

Suppose *S* intends to form a belief about a proposition *p*; as such, on the teleological account at least, *S* intends to accept *p* with the aim₁ of thereby accepting a truth. Deliberation over *whether to believe that p* collapses into the question *whether p is true*. The challenge for the teleologist is to give an account of belief which both explains the presence of Transparency in doxastic deliberation about what to believe, but to do this in a way so as not to exclude non-deliberative beliefs from qualifying as beliefs.

The problem is that the teleologist wants to give an account of belief as being a state brought about through the aim₁ of an agent or some sub-intentional surrogate of aim₁ (an aim₂). If she construes the disposition which constitutes aiming₁ as particularly strong (and one might think that she ought to, if she wants to explain Transparency), then she cannot explain non-deliberative beliefs, which she thinks are aimed₂. If she construes the disposition which

19 Indeed, 'most beliefs are involuntarily formed' and so we should not understand belief aiming at truth 'in such a way that it commits us to supposing that the agent forming the belief has some sort of occurrent intention to form true beliefs whenever a belief is formed' (Pritchard 2011: 244).

constitutes aiming₂ in the non-deliberative case as weak (and one might think she ought to, as non-deliberative beliefs look less strongly regulated for truth), then she cannot explain the Transparency which characterizes deliberative belief formation, when we are aiming₁ at truth. **

2.4.1 Response to the Teleologist's Dilemma

A way to respond to the teleologist's dilemma is to claim that all beliefs share a feature, because they are aim₁ or aim₂ governed. Steglich-Petersen offers a response along these lines arguing that we can capture what it is for a state to count as a belief by an appeal to an agent's intention, that is; an agent 'having as an aim₁ and an interest that the proposition believed is true' (Steglich-Petersen 2006: 514).

The teleologist's dilemma teaches us that this account is inadequate—an appeal to an agent's intention cannot do the work because there are beliefs which do not arise from it but from other cognitive processes. Steglich-Petersen makes the point that whether a cognitive process is weakly regulated for truth is independent of any aim₁ of an agent. This means that we can claim that the cognitive states resulting from non-deliberative belief formation count as beliefs because the characteristic of weak truth regulation that they have is shared by cognitive states which have that very characteristic in virtue of their relation to intentional aims₁ of a believer (Steglich-Petersen 2006: 515).

Steglich-Petersen needs the regulation present in both deliberative and non-deliberative belief formation to be the same—given his claim that it is a feature which the two resulting products share which classifies them all as beliefs. One might think that the regulation present in deliberative truth regulation is strong, however, Steglich-Petersen claims that 'even conscious doxastic deliberation is *de facto* merely weakly regulated for truth' (Steglich-Petersen 2006: 511), it may first-personally feel strong, but third-personally, in actual fact, it is weak, given an agent's epistemic fallibility.

So the claim is that there are cognitive states which share the descriptive characteristic of weak truth regulation with deliberative beliefs, and it is for this reason that they qualify as beliefs on the teleological account. This is a plausible thing to say in some cases; for example, in cases of tacit belief; those beliefs which do not come about through deliberation, but are ascribable to agents, such as my current belief that my house is not on fire. Presumably I have this belief in my cognitive background; though it has not come about through a process of deliberation. It looks though like it is nevertheless regulated for truth, and deliberation on the truth of it (is it the case that my house is on fire?) would yield the same resulting belief (i.e. that it is not). So this might suggest that both deliberative and non-deliberative beliefs are regulated in the same way,

and we might capture this with the claim that they both share the characteristic of *weak truth regulation*.

However, many beliefs are telling against Steglich-Petersen's response, for example, beliefs arising out of transportation, confabulatory explanations of actions, and self-deceptive beliefs.

With respect to the first: Melanie Green and Stuart Brock ran a study investigating the effects of *transportation*, which is 'a mechanism whereby narratives can affect beliefs' (Green and Brock 2000: 701). If a reader is sufficiently engaged in a story, *even if it is presented as a story*, 'they may show effects of the story on their real-world beliefs' (Green and Brock 2000: 701). However, these beliefs are often unstable, that is, they are not retained by subjects. In the study, participants read a story about a little girl who goes shopping with her brother and gets stabbed to death by a psychiatric patient who has escaped from a psychiatric facility. The experiment showed that the more transported participants came away with beliefs in line with the story. So participants might claim to believe that psychiatric patients ought to be kept in higher security facilities, and that shopping malls are dangerous places. Interestingly, these results did not differ between the group in which the story was presented as fact, and the group in which the story was presented as fiction. This looks like a case where regulation for truth by the aim₁ is rather poor.

Perhaps the teleologist might say something like the following about beliefs produced from transportation: when something is presented as fiction and a subject is transported, the subject is still aiming₁ at truth, and so takes it to be the case that the general proposition—in this case some proposition about the dangers of psychiatric patients—is true, and so forms a belief in the proposition. The compelling nature of the story reveals the truth of the general proposition. The subject is still aimed₁ at truth, and her beliefs are weakly regulated for truth.

This is problematic. It might be the case that the subject takes it that the compelling nature of the story reveals the truth of the general proposition, but she should not take it that the story in any way lends support or reveals the truth of that proposition. If the story is presented as fiction, this ought to act as a defeater for the subject taking the story to be expressing propositions. But the experience of transportation is such that this does not occur, which suggests that the regulation for truth in this case of belief formation is weaker than other cases (the non-transportative deliberative case, for example). The subject might *take herself* to be trying to form a true belief, insofar as that is what she thinks she is up to, but the strength of the truth regulation in such a case is weaker than non-transportative cases, and so puts pressure on the claim that the shared feature of weak truth regulation classifies this case as a case of belief.

Now consider confabulatory explanations of one's actions. If a subject suddenly crosses the road upon seeing a black man, or chooses not to hire a female candidate based on a non-gender blind CV, her actions in these cases may be guided by implicit biases against certain racial groups or women. When asked about her behaviour, she might confabulate by claiming, in the first case, that the man looked threatening, or that she needed to cross the road because she mistakenly thought she saw a friend on the other side, and in the second case she might claim that the CV of the candidate really was not good enough for the job. Here she offers confabulatory explanations of her actions, and let us suppose that she *believes* the explanations she offers. Because she is motivated to think of herself as person of egalitarian persuasion, who does not have racial or gender prejudices, she forms a belief to explain her behaviour which is influenced by motivational considerations. I am not saying that these cases are problematic for the teleological account because they are cases in which the subjects do not intend to believe truths. Though their motivations might make them poor believers of truths, this could still be their intention, indeed, in such cases subjects may be seeking to form a true belief, just as they do in other cases. Rather my claim is that the truth regulation present in these cases is sufficiently different from the truth regulation present in cases where motivational factors are not playing a role. If the subject were *explicitly* racist or sexist, and not ashamed by this, and she acted in the above ways, and was asked to explain her behaviour, we might expect her to say that she crossed the road to avoid the black man because black men are dangerous, or she did not hire the woman because women make for bad colleagues. In these cases, the strength of truth regulation is less affected by motivational attitudes. Once the subject is motivated to think of herself as someone who does not have racist or sexist attitudes, or motivated for other people to not think of her in that way, the regulation for truth in her belief formation is weakened.

Looking now to self-deception, it is agreed upon in the literature on self-deception that beliefs resulting thereof include a motivational element which can involve the misreading or ignoring of evidence in coming to a belief. This might be a pro-attitude towards a proposition being *true* (wishful self deception), for a proposition being *believed* (wilful self-deception), or for a proposition being *false* (dreadful self-deception) (van Leeuwen 2007a: 423–5). Take the case of the cuckolded husband who self-deceptively believes that his wife is faithful. Evidence that she is unfaithful may be available to him (insofar as it is perceptually available—he sees that his wife returns home late, dishevelled, and uninterested in him), but an alternative cognition, such as the belief that his wife is having an affair, is motivationally unavailable, due to the husband's very strong motivation for it to be the case that his wife is faith (wishful self-deception) or for it to be the case that he *believes* that his wife is faithful (wilful self-deception). Again, my claim here is not

that the self-deceived subject is not trying to believe truly, rather, it is that what constitutes that trying is sufficiently different from what constitutes it in the deliberative case where pro-attitudes are not playing a significant role in the fixation of belief. We might understand this difference by claiming that in these cases the agent is aiming₁ at the truth, that is, she takes herself to be trying to form a true belief, but she is not aiming₂ at the truth, that is, sub-intentionally, something else is going on, and so the truth regulation that constitutes aiming₁ is significantly different from cases of aiming₁ in the absence of pro-attitudes towards the target proposition.

We can see already then the beginnings of a distinction between beliefs like these (arising from transportation, confabulation, and self-deception) and deliberative beliefs resulting from transparency to considerations of truth. Let us call this first set *bad beliefs* in virtue of their being less epistemically good, and being influenced by non-epistemic factors. A proponent of the teleological account will claim that all of these beliefs share some feature grounded upon their being governed by an aim₁ or aim₂. Now, it is of course no argument against this that bad beliefs are often false—this is consistent with truth regulation being weak. What is important though, is that the fixation of bad beliefs is influenced by non-epistemic factors, which makes it such that the truth regulation present in these cases is weaker than the truth regulation present in cases in which one does not hold a pro-attitude towards a proposition's being true.

In the case of transportation, for example, we can compare to a case in which we are not transported, in which we deliberate, and would be wary of taking a fiction as an evidence source for a belief about the dangers of psychiatric patients. Even if beliefs arising out of transportation are deliberative beliefs—so that having seen the fiction, we ask somebody the question of whether they believe psychiatric patients present a danger, and they deliberate, and reasons for them being a danger are salient and so they form a belief—this is still a problem for the teleological account because the transformative experience has influenced the product of deliberation and produced a belief which is less strongly regulated for truth. In the case of confabulation, my desire to be a person of egalitarian persuasion, or at the very least, to be viewed that way by my peers, influences my belief formation such that I form beliefs about my actions which are not regulated for truth in the way they would be if I did not have the desire to be a supporter of equality, or at least, be perceived as one. In cases of self-deception, the pro-attitude towards the target proposition plays a role in generating a belief in that proposition, a belief which would not be acquired were the pro-attitude absent.

One worry about my strategy here is that I am pointing only to a difference in regulation in degree, and not in kind, and this is consistent with Steglich-Petersen's characterisation of the

truth regulation being ‘weak’ in both cases. But Steglich-Petersen himself is at pains to point out at the start of his response that the regulation present in the deliberative case is *not* strong, objectively speaking, it is weak, a characteristic it shares with non-deliberative belief formation (Steglich-Petersen 2006: 511). If a difference in degree were not enough to make trouble for Steglich-Petersen, we would not expect him to worry about it when he articulates his response. But he does worry about it and so I take it that a difference in degree between the regulation present in cases of deliberative belief without the motivational element, and cases in which pro-attitudes influence the fixation of deliberative belief, is sufficient to cause problems for the teleological account. So though we can accept that regulation for truth in deliberation is weaker than it feels first personally, it is clearly different from the truth regulation present in cases of what I have identified as bad belief.

To be clear; it is not my argument that bad beliefs are not beliefs, I think that they are. Rather, my claim is that it is implausible to suggest that the reason bad beliefs are beliefs is because they share some regulatory feature with other beliefs. If regulation of different strengths goes on, and it is by appeal to this feature that we classify beliefs, then why count bad beliefs as beliefs? Bad beliefs do not present a problem for my preferred account, because I demarcate belief from other cognitions by appeal to their motivational role, as defended in chapter one. Providing these beliefs play the motivational role of belief as specified in the last chapter, they qualify as beliefs on my account. It is motivational role we ought to focus on then, and not truth regulation, as it is the former which can bring together all those states, including bad beliefs, which are beliefs.

To sum up this section: the teleologist’s dilemma challenged the teleologist to give a descriptive account of belief which both explained the Transparency and did not exclude non-deliberative beliefs from counting as beliefs. In response, Steglich-Petersen claimed that third-personally, due to our epistemic fallibility, even beliefs from deliberation are only weakly regulated for truth, and that this feature of weak truth regulation is shared by non-deliberative beliefs. Further, the reason we get to call these non-deliberative states beliefs on teleological grounds is because they share the feature of weak truth regulation with deliberative cognitive states which have that feature in virtue of their relation to intentional aims of a believer.

I claimed that this line might be plausible in some cases of non-deliberative beliefs, but suggested that beliefs resulting from, for example, transportation, confabulation, and self-deception cause a problem for Steglich-Petersen’s response. This is because non-epistemic factors influence the deliberative fixation of belief, and this indicates a different strength of truth

regulation. Hence it is difficult to group all beliefs together by appeal to the claim that they share some regulatory feature. I claim then that the teleologist's dilemma remains unanswered.

2.5 Conclusion

In this chapter I outlined the teleological account and its explanations of the three features of belief outlined in chapter one. I argued that the explanations offered are problematic for two reasons. Firstly, the aim_1 of belief as a *cause* of some phenomena does not sit comfortably in a causal story once we distinguish between triggering and structuring causes. Second, if belief had an aim_1 , this ought to be weighable, and if it were weighable, it would not be at the required strength to explain any of the three features. I argued for the claim that the so-called ' aim_1 ' of belief is not weighable by considering Owens's Exclusivity Objection. I considered two responses: one which accepted exclusivity, and one which denied it. I argued that these responses did not work because motivational factors do not play a role in affecting what we believe by providing non-epistemic reasons for belief, but rather by changing the confidence levels required for belief.

I also considered the teleologist's dilemma. One response to this is to appeal to weak truth regulation as a feature which is shared by both deliberative and non-deliberative beliefs. I argued that some beliefs put pressure on the claim that there is a basic level of truth regulation common to all beliefs, and when there is variety, we might question why all of the cases qualify as cases of belief on the teleological account.

In conclusion, the teleological account does not work independently (insofar as it does not provide the resources for distinguishing belief from other attitudes and capturing all beliefs *as* beliefs), it cannot explain any of the three features which indicate a connection between belief and truth, and two objections to it remain unanswered. Therefore, the teleological account ought to be rejected.

3

THE NORMATIVE ACCOUNT OF BELIEF

In this chapter I will consider the normative account of belief and argue that the best version of it is one which claims that belief is governed by a norm couched in permissibility, instead of obligation. I will then discuss how this kind of account might do the explanatory work required of the belief theorist in accommodating the three features outlined in chapter one. It is the account cast in these terms, which poses an alternative to my account which will be developed in chapters five and six.

In the account of the connection between belief and truth I will formulate, I argue that belief is not constitutively or conceptually normative, and that a weaker version of Epistemic Normativity should instead be understood by appeal to the standards laid down by biology. Further, I argue that a stronger Epistemic Normativity should not be understood as a feature linking belief to truth, but rather as one present due to our mechanisms for belief-production performing their function of producing useful beliefs (though not useful as an approximation to truth). The thought is that we make a mistake with respect to the strength of Epistemic Normativity, biological norms lay down a standard from which beliefs can deviate, and it is by reference to this standard that we can explain a weaker version of Epistemic Normativity, and we can explain the illusion of a stronger Epistemic Normativity by appeal to our mechanisms for belief-production performing their function of producing *useful* beliefs.

The competitor to my approach with respect to Epistemic Normativity is one which appeals to *sui generis* epistemic norms, which are irreducible to biological norms, and takes belief's connection to truth to be an 'internal, normative one' (Shah 2003: 454). I will be concerned with whether the most plausible version of this view can do the explanatory work called for by the three features of belief outlined in chapter one, and whether, in virtue of taking a stronger view about the nature of Epistemic Normativity, it can do work which my account, with its reducing Epistemic Normativity to biological normativity, cannot.

I will consider the No Guidance Objection to the normative account and suggest that it can be answered. However, the reply to the objection shows that just as the teleological account makes a mistake in grounding the connection between belief and truth at an intentional level, and citing only a triggering cause for the phenomena—so too does the normative account make this mistake. I will conclude with the claim that the most plausible version of the normative account cannot give an explanation of the three features of belief outlined in chapter one, and so it ought to be rejected.

3.1 The Normative Account

In the last chapter we saw that the teleological account of belief could not respond to objections to it, and could not account for the connection between belief and truth, and so it ought to be rejected. Normative accounts of belief offer an alternative.

Normative theorists claim that belief is norm governed and this is a conceptual truth about belief (though it need not be, one could be a normativist whilst taking the claim to be constitutive of belief but not conceptual). However, there has been considerable debate over what the norm governing belief is, and what explanatory work it can do. Where normative theorists agree is on the claim that belief is governed by a norm, and this entails several other claims. One is that understanding the concept of belief requires understanding that a belief is correct (if and) only if it is true:²⁰

Marco could not be said to understand what it is for Ebenezer to *believe* that Mallory reached the summit unless he understands that Ebenezer's belief is correct only if Mallory did reach the summit, unless he understands, that is, that Ebenezer ought to believe that Mallory reached the summit only if he did. (Boghossian 2003: 38)

It is part of our concept of belief then, that a belief is correct (if and) only if it is true. The norm governing belief 'is a constitutive and conceptual norm, which a subject needs to possess to be able to use the concept of belief' (Engel 2013: 203). Given this, belief attributions are also normative, our being able to ascribe belief also depends on our understanding that a belief is correct (if and) only if it is true. Normativists claim that belief's being conceptually or constitutively normative serves to distinguish it from other cognitive states, such as considering *p*, supposing *p*, or imagining *p*. In these cases it is not the case that one ought to hold the attitude toward *p* only if *p* is true (Gluer and Wikforss 2009: 42).

Shah, for example, claims that belief's connection to truth is one which is internal and normative (Shah 2003: 454). I deny this, my claim is that belief's connection to truth is contingent and can be explained by appeal to biology. With respect to belief's standard of correctness, the normativist claims that truth is the standard of correctness for belief, and that this is a conceptual matter, such that one does not understand the concept of belief, unless one understands what it is to take belief's standard of correctness to be truth (Shah 2003: 466).

²⁰ Some normative theorists formulate the norm governing belief as a biconditional. Objections have been raised to this formulation, as we will see in §3.2

Belief's having this standard of correctness, and this being a conceptual matter, means that 'a competent user of the concept of belief must accept the prescription to believe that p only if p is true for any activity that he conceives of as belief-formation' (Shah 2003: 470). My preferred account can generate the correctness condition by appeal to the standards laid down by the biological history of our mechanisms for belief-production.

It is worth spelling out what I mean by *biological normativity*. First note that the standard of correctness and norms regarding how to meet that standard which are generated from biology are not ones which involve prescription, from a biological norm nothing follows about what we ought to do, as David Papineau notes it 'is a vulgar, and indeed dangerous, error to infer, from the premise that X has been biologically designed to do Y, that in some sense X *ought* to Y' (Papineau 1999: 21, fn. 5). Rather, some item's having the function to F means that there is something it has been designed to do, and it is only in a very weak sense, that biology therefore lays down a standard from which a token trait can deviate. So though our mechanisms for belief-production have the function of producing true beliefs, that is not to say that we are in any sense *obligated* to believe truly. On my account then, the correctness condition for belief can be generated by appeal to what our mechanisms for belief-production have been designed to do, which explains why they exist with the features they have.

There is a question then of whether the correctness condition of belief roots itself in a norm which is not reducible to biology, and is somehow related to obligation—this is a claim we can draw out from the normative account. We will see towards the end of the chapter, that giving the correctness condition of belief this second reading makes it difficult to give an account of the connection between belief and truth. The difficulty is the analogue of the difficulty faced by the teleological account—that of rooting the connection to truth at the intentional level, rather than at the sub-intentional level secured by biology.

As we saw in chapter one (§1.3.1), Shah and Velleman argue that if we are interested in demarcating belief from other attitudes, this can be done by appeal to its connection to truth. They argue that belief can not be demarcated by appeal to its motivational role, because this role is shared with other attitudes. They cite Bratman's (1992) examples of acceptance in a context as showing that there is at least one other cognitive state which shares the motivational role of belief, and Velleman's (2000) arguments as demonstrating the stronger claim that belief's supposed motivational role is shared by 'all the cognitive attitudes, including imagining that p ' (Shah and Velleman 2005: 530, n. 3).

I argued in chapter one that these arguments against demarcating belief on the basis of motivational role do not work, because non-belief attitudes can play the role only in conjunction

with some other attitudes, and only in some contexts, whereas belief plays the motivational role *by itself*, and *across contexts* (§1.2). Though I also claimed that the motivational account could not give an explanation of the three features of belief which indicate a link between it and truth (§1.4), so something else is required. What can do this work, Shah and Velleman think, is the way in which belief is regulated, that is, the way in which beliefs are ‘formed, revised, and extinguished’ (Shah and Velleman 2005: 498). Beliefs, it is claimed, are regulated for truth, other cognitive attitudes—imaginings, supposings, considerings—are not. It is their being so regulated which is part of the concept of belief, when we conceive of an attitude as a belief, we conceive of an attitude which has been regulated for truth (Shah and Velleman 2005: 498). Being regulated for truth though does not exhaust the concept of belief, also included is belief’s standard of correctness. This means that conceiving of an attitude as a belief means conceiving of an attitude as being correct (if and) only if it is true. The concept of belief then is ‘that of a cognition that is governed, both normatively and descriptively, by the standard of truth’ (Shah and Velleman 2005: 498). It is governed descriptively by a standard of truth, because it is a state which is at least weakly regulated for truth. It is governed normatively by a standard of truth because it is a state which is constitutively norm governed such that a belief is correct ‘if and only if [it] is true’ (Shah and Velleman 2005: 501), where this normative governance is not reducible to biological normativity. Much of this is granted by my preferred account. It is certainly right that beliefs descriptively have a standard of truth, in virtue of the mechanisms which produce them having the relational proper function to produce true beliefs (I will explain what this means in §5.3). However, my account claims that belief normatively having a standard of truth is right, but only insofar as the biological history of the mechanisms which produce belief lay down a standard from which the products of those mechanisms can deviate (I have moved from talking of governance by a standard to merely *having* a standard, because of concerns I will raise in §3.4). The normativity of belief need only be understood as fixed by the biological history of our mechanisms for belief-production.

3.2 The Too Demanding Objection

In what remains of this chapter I will suggest that the most plausible version of the normative account is one cast in terms of permissiveness. First I will consider an objection from Krister Bykvist and Anandi Hattiangadi, and discuss some responses to it which involve a restatement of the norm governing belief. In the end, I will suggest that the objection shows that the norm of belief ought to be formulated in terms of *permissibility*. I will take this forward in the remainder of the chapter and see whether the permissive normative account can explain the three features

of belief outlined in chapter one, and thus whether the account represents a genuine alternative to my view.

An objection frequently raised against the biconditional formulation of the norm of truth governing belief which is found in Shah (2003)²¹ is that once it is broken down into two conditionals, one of those conditionals is too demanding. Bykvist and Hattiangadi state the standard of correctness for belief as:

(1) For any p : the belief that p is correct if and only if p is true. (Bykvist and Hattiangadi 2007: 277)

Interpreting 'correct' as a normative term, Bykvist and Hattiangadi state the norm of truth governing belief as:

(2) For any S, p : S ought to believe that p if and only if p is true. (Bykvist and Hattiangadi 2007: 277)

The narrow scope reading of (2) is:

(2a) For any S, p : S ought to (believe that p) if and only if p is true. (Bykvist and Hattiangadi 2007: 278)

If we pull out the two conditionals we get:

(2a*) For any S, p : if the proposition that p is true, then S ought to (believe that p).

(2a**) For any S, p : if S ought to (believe that p), then the proposition that p is true. (Bykvist and Hattiangadi 2007: 279)

The Too Demanding Objection is based on conditional (2a*). Bykvist and Hattiangadi claim that in its being too demanding, it is false. This is because it places too much strain on our relatively limited cognitive faculties, that is, it makes it the case that one ought to believe *every* true proposition. As the number of true propositions available to be believed is infinite, and indeed there are propositions too complex for us to believe (for example, huge conjunctions of true propositions), the requirement in (2a*) can not be met. As Bykvist and Hattiangadi put it:

²¹ The same formulation is found in (Shah and Velleman 2005), but Bykvist and Hattiangadi do not cite this work in their paper, and so I will assume that they did not have Shah and Velleman's paper in mind as their target.

it is not *humanly* possible to believe arbitrarily complex propositions [...] Limits on how much information a human brain can store, and limits on the length of human lives place limits on the complexity of the propositions we can believe. (Bykvist and Hattiangadi 2007: 279)

But, Bykvist and Hattiangadi claim, ‘ought’ implies ‘can’, and so (2a*) is false (Bykvist and Hattiangadi 2007: 279).

This worry has been raised elsewhere, Engel argues that the norm of belief cannot be formulated as the claim (2a*) because ‘there are plenty of propositions which are true, and which are not worth believing, because we cannot clutter our minds with trivialities’ (Engel 2006: 368). The objection shows not that there cannot be a norm for belief, but that we need to formulate it in a different way (Papineau 2013: 66). I will consider some possibilities for a reformulation below. I will claim that casting the norm in terms of permissibility is the way to answer the Too Demanding Objection, but we will see that though the account is then no longer vulnerable to the Too Demanding Objection, it is problematic in its explanation of the three features outlined in chapter one.

3.2.1 Believe Only True Things

The very least that we should take from Bykvist and Hattiangadi’s argument—if not that the normative account fails—is that the norm of belief needs to be otherwise formulated so as not to be too demanding. Paul Boghossian pre-empts this kind of worry when formulating his normative account. He claims that ‘no one thinks that it’s a norm on belief that one believe *everything* that’s true’ (Boghossian 2003: 37). Rather, we should understand the norm of belief as:

[(3)] For any p : One ought to believe that p only if p . (Boghossian 2003: 37)

This norm requires that one believe only true things. However, Bykvist and Hattiangadi claim that the move from a norm like (2) to a norm like (3) does not help the normativist. This is because this new formulation is not normative, it does not put the subject under any obligation in any circumstances. (3) is not instructive; from it we get no suggestion of what S is obligated to believe. Let us take the case in which p is true and the case in which p is false to demonstrate this. If p is true we get the result that S is under no obligation to believe that it is so. More interestingly, if p is false, it is not the case that S ought *not* to believe p . Rather, all we get from (3) is that it is not the case that S ought to believe that p . There is an important difference: all we can derive from ‘it is not the case that S ought to believe that p ’ is that S does not have an

obligation to believe that p . This is different from what we get from ‘ S ought not to believe that p ’, from this we derive that S has an obligation *not* to believe that p . So the former, which is what we draw from (3) is compatible with it being permissible for S to believe that p . So *whatever* the truth value of p , from (3) we do not get any indication of what S ought to believe (Bykvist and Hattiangadi 2007: 280). The same point is made by Kathrin Gluer and Asa Wikforss when they claim that in rejecting the stronger norm captured in (2), ‘Boghossian falls back on a norm that is much too weak to serve a guiding function’ (Gluer and Wikforss 2009: 43). If it is to do the explanatory work in explaining why belief has the features it does, the norm needs to be able to serve this function (see §3.4).

3.2.2 Propositions We Consider

Another formulation of the truth norm which might look to meet the Too Demanding Objection is one cast in terms only of those propositions we *consider*. Ralph Wedgwood offers this norm claiming that ‘the fundamental epistemic norm [...] only concerns propositions that one *actually consciously considers*’ (Wedgwood 2002: 273). Bykvist and Hattiangadi capture Wedgwood’s reformulation as:

[(4)] For any S, p : if S considers whether p , then S ought to (believe that p) if and only if p is true.
(Bykvist and Hattiangadi 2007: 280)

Wedgwood’s reformulation of the truth norm as applying only to those propositions a subject consciously considers answers the Too Demanding Objection. For now, one only ought to believe that p if and only if p is true, but only for those propositions considered. This means that we do not get the unhappy result that believers ought to believe *every* true proposition. If I do not *consider* some infinitely complex proposition, then the truth norm puts me under no obligation to believe it. Rather, if I consider whether p , *then* I ought to believe that p if and only if p is true.²²

However, a related problem arises for Wedgwood’s reformulation. Bykvist and Hattiangadi, following Roy Sorensen (1988), call ‘blindspots’ those propositions for which it ‘is logically impossible to believe them truly: if they are true, then you don’t believe them, and if

²² However, this norm will cause problems for those whose attention wanders. If it is pointed out to me that the number of blades of grass in the garden might be odd, in your bringing this to my attention, if I consciously consider it, even for a moment, I am obligated to find out the truth with respect to this proposition (thanks to Paul Noordhof for pointing out this consequence).

you believe them, then they are false' (Bykvist and Hattiangadi 2007: 281). They give the following two examples of blindspot propositions: '[i]t is raining and nobody believes that it is raining' and '[t]here are no believers' (Bykvist and Hattiangadi 2007: 281), let us label these propositions (a) and (b) respectively. Now, if (a) is true, it follows that the subject does not believe that it is raining, if the subject believes (a), it follows that (a) is false. Equally, if (b) is true, it is not believable, if it is believed, then it is false (Bykvist and Hattiangadi 2007: 281). ****

Now let us—as Bykvist and Hattiangadi do—take proposition (a) and insert it into Wedgwood's preferred formulation of the truth norm. This is the result:

[($4_{\text{blindspot}}$)] For any S , p : if S has considered whether it is raining and nobody believes that it is raining, then S ought to (believe that it is raining and nobody believes that it is raining) if and only if it is raining and nobody believes that it is raining is true. (Bykvist and Hattiangadi 2007: 281)

As we have seen, proposition (a), if true, implies that the subject does not believe it. If the subject does believe (a), then it would be false. But looking at ($4_{\text{blindspot}}$), we are told that the subject ought to believe (a) if and only if it is true. This is problematic because (a)'s being true is incompatible with the subject believing (a) to be true. With respect to blindspots, Wedgwood's reformulation of the truth norm in terms of propositions one consciously considers does not help the normativist. The problem here is that once again, we have a violation of ought implies can—the obligation put on a subject by the truth norm, in cases of blindspot propositions, cannot be satisfied.

Blindspot propositions are going to be problematic for any formulation of the norm governing belief in terms of truth, expressed as an *ought*. Any norm which prescribes believing that p or not on the basis of whether p is true will be faced with violating the *ought implies can* principle. The problem is especially important when discussing Wedgwood's formulation because though other violations of the *ought implies can* principle with regards to propositions too complex to believe are solved by restricting the norm to propositions one considers, blindspot propositions are ones which can be considered, but cannot be believed truly.

Compare with my account which takes belief's standard of correctness and norms derived from that regarding how best to meet it, to be laid down by biology. Epistemic Normativity is reducible to biological normativity and the latter does not produce obligation. This means that blindspot propositions are not problematic for my account, because I give a reductive account of Epistemic Normativity grounded in biology, and argue that biological norms are not ones which explicitly guide an agent's activity—as noted above, biological norms are not prescriptive. This means that blindspot propositions are ones which cannot be believed truly, but that is not a

problem for my account because I do not explain the connection between belief and truth by appeal to a norm which guides belief formation, I appeal instead to one of the functions of the mechanisms which produce beliefs.

3.2.3 Permissiveness

Let us look back to Shah's formulation of the truth norm. Bykvist and Hattiangadi were not targeting Shah and Velleman's formulation (see fn. 21), which is the same as Shah's. In Shah and Velleman's paper, the Too Demanding Objection is pre-empted. Before replying to objections they foresee to their account, Shah and Velleman remind us that belief's standard of correctness is captured by the biconditional norm 'a belief is correct if and only if it is true' (Shah and Velleman 2005: 519). This norm, also offered by Shah (2003: 473), once formulated by Bykvist and Hattiangadi as (2), falls foul of the Too Demanding Objection. However, Shah and Velleman immediately go on to say that correctness is not an injunctive notion, but a permissive one. What this means is that the norm forbids a subject from believing that p when p is false, but it only *permits* a subject to believe that p when p is true. As Shah and Velleman note: '[o]ne is not required to hold every belief that would be correct' (Shah and Velleman 2005: 519). Rather, once one enters into deliberation regarding p , 'one is committed to forming the belief if it would be correct, and this commitment tends to supply the injunctive half of a biconditional norm, mandating a belief in p if and only if p is true' (Shah and Velleman 2005: 519).

However, Daniel Whiting responds to the Too Demanding Objection by reformulating the norm of truth in terms of *permissiveness*. He cites Shah and Velleman's formulation as problematic *given* the Too Demanding Objection (Whiting 2010: 218, fn. 8) and so does not take Shah and Velleman's noting that correctness is permissive and not injunctive, as a satisfactory pre-emption of the Too Demanding Objection, perhaps because though correctness can be understood permissibly, a norm featuring an 'ought' cannot be. He claims that Bykvist and Hattiangadi's way of cashing out the normativity contained within (2) ('[f]or any S , p : S ought to believe that p if and only if p is true' (Bykvist and Hattiangadi 2007: 277)), is inappropriate. It is the inappropriateness of their formulation of the normativist's thesis that gives their objection any force. Whiting claims that not all normative uses of 'correct' express an 'ought', some express merely a 'may'. If we take the normative thesis to be an example of this use of 'correct' then it does not fall foul to the problem Bykvist and Hattiangadi claim for (2a*).

Whiting offers some examples of cases in which the normativity involved in the use of 'correct' expresses not an 'ought' but a 'may'. One example comes from chess and the correct way to move the rook. Whiting claims that to say that it is correct when playing chess to move a

rook horizontally is not to say that one *ought* to move it horizontally. This is because one may be better off moving it in another permissible way, or indeed, by not moving it at all (Whiting 2010: 215). All that one should take from the statement that it is correct to move a rook horizontally is that it is permissible to do so, that is; one *may* do so, not that one *ought* to do so (Whiting 2010: 215).

Whiting's response consists of two claims. Firstly, to say that an act is correct is to say that one may perform it, not that one ought to. Secondly, to say that an act is incorrect is to say that one ought not to perform it (Whiting 2010: 215). Clearly it is the first of these claims which is doing the work for Whiting as it supports the claim that he needs in responding to Bykvist and Hattiangadi, that is, 'what is correct does not have immediate consequences for what one ought to do' (Whiting 2010: 215). Shah and Vellman might agree with this, as they say that correctness is to be understood permissibly, not injunctively. But as noted above, though *correctness* can be understood this way, a norm featuring 'ought' cannot be. Whiting's move is to have the permissibility of correctness featuring in the norm of belief by changing 'ought' to 'may'.

From this, Whiting suggests an alternative way to that of Bykvist and Hattiangadi of bringing out the normativity of (2):

[(5)] For any S, p : S may (believe that p) if and only if it is true that p . (Whiting 2010: 216)

This breaks down into:

[(5*)] For any S, p : if it is true that p , then S may (believe that p)

[(5**)] For any S, p : if S may (believe that p) then it is true that p . (Whiting 2010: 216)

Whiting goes on to argue why this reformulation does not fall foul to the problems we saw with (2a*) and (2a**), and so claims that he has vindicated the normative account.

A worry at this point might be that just as (2a*) was in trouble when coupled with the thesis that *ought implies can* because it would require too much of believers, (5*) faces a similar problem. There is still a potentially infinite number of true propositions, and (5*) makes it permissible for one to believe those propositions. Whiting grants that this may sound strange—how could it be the case that one is permitted to do something which one is unable to do, that is; how could one be permitted to believe all of these complex truths when one is not able to do so? However, Whiting claims that this should not concern us, because (5*) is equivalent to:

([5***]): For any S, p : if it is true that p , then it is not the case that S ought not to (believe that p).
(Whiting 2010: 217)

Whiting claims that we should not be worried by the fact that for any given complex or trivial proposition, it is not that case that we ought not to believe it (Whiting 2010: 217).

The Too Demanding Objection then is answered by an appeal to permissiveness. We can understand correctness as permissive, but we cannot then have a norm which prescribes what one *ought* to do. We need to change the ‘ought’ to a ‘may’ in the norm governing belief. Construing the norm as merely permissive is also claimed to meet the worry raised by blindspot propositions, those propositions for which ‘it is logically impossible to believe them truly: if they are true, then you don’t believe them, and if you believe them, then they are false’ (Bykvist and Hattiangadi 2007: 281). Look again at ‘it is raining and nobody believes that it is raining’. If that proposition is true, on the *ought* version of the normative account, it follows that the subject does not believe that it is raining, but that the subject ought to believe the proposition, but if she does then the proposition is false. On the permissive account though, if the proposition ‘it is raining and nobody believes that it is raining’ is true, the subject is merely permitted to believe it. We might worry that being permitted to do something impossible is a bad result, but as we just saw, Whiting covers this problem in his discussion of being permitted to believe all truths. Perhaps there is room for a residual worry here concerning the permissibility proponent’s treatment of blindspot propositions. If a blindspot proposition is true, we are not obligated to believe it, but merely permitted to. But how can we be permitted to do something which, if we did it, would be a mistake, given that blindspot propositions cannot be believed truly, and I am only permitted to believe true propositions? It is not obvious then that the problem of blindspot propositions *is* dealt with by moving to permissibility after all. However, let us put this worry aside—if the normative account cast in terms of permissibility can do the explanatory work with respect to the three features of belief outlined in chapter one, but gets a bad result only in these strange blindspot cases, we might want to keep hold of it. Let us turn then to look at whether or not the normative account cast in terms of permissibility *can* do the explanatory work called for by the three features of belief which indicate a link between it and truth.

3.3 Explanatory Work

By looking at the Too Demanding Objection we have seen that the norm of truth must be understood as permissive. This is the formulation of the normative account which I will take to be a competitor to my approach as developed in chapters five and six.

Explanations of Transparency, the Uncontrollability Thesis, and Epistemic Normativity have been offered by normativists who construe the norm governing belief in terms of obligation. I will outline how the permissibility theorist might explain these features. I will suggest that only some of these explanations offered by both accounts, are successful. I will argue that though the account cast in terms of permissibility avoids the Too Demanding Objection (though it is still up for grabs the extent to which it avoids the problem raised by blindspot propositions), it is unable to explain all three of the features of belief outlined in chapter one.

Looking first at Transparency, what needs explaining is why deliberation over *whether to believe that p* is answered only by considerations relevant to the question *whether p is true* (Shah 2003: 467). Later (§3.4.3), I will raise similar worries to the explanation of Transparency available to the normativist as I did against the explanation offered by the teleological account in the last chapter (§2.2). The problem is that the role of truth given by the normative account for the three features of belief which indicate a link between it and truth, is given at the intentional level, at the level of believers recognising a norm governing belief, rather at the sub-intentional level as secured by biology. This is problematic when we look at the explanations offered for the three features in terms of triggering and structuring causes. For now though, my claim will be that Transparency can only be explained by appeal to a norm, if that norm is expressed as a biconditional, that is, if the norm requires the believer to believe *all truths*. The explanation offered by Shah is the following: when an agent frames her deliberation with the question *whether to believe that p*, two dispositions are activated: one to be moved by considerations which are, by her lights, relevant to the truth of *p*, and one which blocks considerations which are, by her lights, irrelevant to the truth of *p*. It is part of the understanding the concept of belief that one is disposed in such ways when framing one's reasoning with the concept of belief (Shah 2003: 467). If the norm governing belief formation were that I ought to believe that *p if and only if p*, then, once I recognise that *p* in deliberation, I move immediately to belief that *p*. This norm being in play would explain Transparency. But we have seen this cannot be the norm in play because 'no one thinks that it's a norm on belief that one believe everything that's true' (Boghossian 2003: 37), a norm formulated this way would fall foul to the Too Demanding Objection (§3.2). Rather, when a subject exercises the concept of belief, she accepts that she ought to believe that *p only if p* is true (Shah 2003: 449), and if it is this which she recognises, then her recognition of this cannot explain Transparency.

If the subject accepts that she ought to believe that *p only if p* is true, then she is forbidden from believing that *p* when *p* is false, but only permitted to believe *p* when *p* is true. It is

consistent with the subject recognising this, that she could have all the reasons for p before her, she could answer the question *whether p is true*, recognise that there is sufficient evidence to believe that p , and not believe that p . After all, she recognises that she ought to believe that p *only if p* , not *if and only if p* . But this is not possible in doxastic deliberation. When we deliberate over *whether to believe that p* , we move immediately to the question *whether p is true*, and answering this second question answers the first. We are moved from *whether p is true* to *belief that p* , and this cannot be explained by a norm governing belief, unless that norm is formulated in a way which falls foul to the Too Demanding Objection.

An explanation of Transparency is also not available to the permissibility theorist, for the same reason. If when I exercise the concept of belief this involves accepting that I *may* believe that p if and only if p is true, this may too activate the disposition to be moved only by considerations relevant to the truth of p and the disposition to block those considerations which are irrelevant to the truth of p . When putting to myself the question *whether to believe that p* , I recognise that I am permitted to believe that p if only if p is true, and so I immediately move to the question *whether p* . However, we saw in chapter one (§1.3.1) that the phenomenon of Transparency is the *collapsing* of one question into another. When I answer the question *whether p is true*, I too have answered the question *whether to believe that p* . There is no inference, there is no need for a bridge premise which moves me from *whether p is true* to believing that p . Transparency cannot be accounted for by appeal to a permissibility norm governing belief because if I have all the evidence for p before me, and I answer the question *whether p is true* in the positive, I recognise that it is *permissible* to believe that p , not that I *ought* to believe that p . Why think that mere permissibility would move me from an answer to *whether p is true* in the positive, to believing that p ? If I have all the evidence that p is true, but then do not believe it, this is consistent with my belief formation being governed by the norm that I may believe that p if and only if p . If p is true, I am not obligated to believe that p (I merely *may* believe it, from (5*)). We cannot claim that once in the game of deliberating over *whether to believe that p* , one is committed to forming a belief with respect to *whether p is true*, and so will immediately move from p to *belief that p* . This is because *even if* I am committed to forming a belief with respect to whether p , answering the question *whether p is true* in the positive is compatible with my failing to believe it, because I am merely permitted to do so. Of course if I am committed to forming a belief on the matter, then I might, *as a matter of fact*, move from *whether p is true* to *believing that p* . But if the norm governing my belief formation is merely permissive, I should be able to answer *whether p is true* in the positive, and yet feel no obligation to believe that p , and so fail to believe that p .

The obligation account formulated as an *only if* and the permissibility account which formulates the norm as a *may*, face the same problem with respect to explaining Transparency. This is that the norm both accounts appeal to is not at the required strength to explain Transparency, because Transparency ensures that I cannot answer *whether p is true* in the positive and fail to believe that *p*. If the norm governing belief were a norm cast in terms of an *only if* as with the obligation theorist, or in terms of a *may* as with the permissibility theorist, then I ought to be able to do this. However, if the normative theorist goes to a strength of norm to explain Transparency, (an *if and only if* norm), she is then back in trouble with respect to the Too Demanding Objection.

Let us look now to the Uncontrollability Thesis, which is the impossibility of ‘unmediated conscious belief-production’ (Noordhof 2001: 248). Normativists have explained the Uncontrollability Thesis by appeal to Transparency:

due to [T]ransparency, considerations that are recognized within doxastic deliberation as relevant to determining the truth with respect to *p* must be immediately seen as relevant to determining *whether to believe that p*. (Shah 2003: 453)

Considerations being seen as relevant to determining *whether p is true* might be seen as relevant to determining *whether to believe that p*, but we have seen that the questions should not collapse into one another if the norm guiding doxastic deliberation is formulated as *only if*, and this means that the normativist cannot explain Transparency. The worry with appealing to Transparency to explain the Uncontrollability Thesis, or at least, citing Transparency in an explanation of the Uncontrollability Thesis, is that the normativist simply pushes the explanatory project back a step—now we ask her to explain Transparency—and then she has no explanation, as we saw above. It would be much better for the normativist if she could explain the Uncontrollability Thesis *without* appealing to Transparency, because the latter is something which she does not have the tools to explain. I think that this can be done, as I will argue below after dismissing the explanation of the Uncontrollability Thesis which appeals to Transparency.

Shah claims that the possibility of a connection between judgements about what it might be practically good or advantageous to believe and a resulting belief, is blocked by Transparency (Shah 2003: 453). If I judge that it would be good for me to believe that *p*, this judgement, as one not relevant to the truth value of *p*, will thus not be a consideration permitted by Transparency. Non-truth related considerations are ‘ineligible to play an immediate role in doxastic deliberation’ (Shah 2003: 453). A subject cannot believe that *p* as a result of having the conscious intention to believe that *p*, without there being a mediating act to help produce the

belief, or evidential support for p . Shah claims that it is a conceptual impossibility that pragmatic considerations alone could be reasons to believe (Shah 2003: 453), and so without a mediating act to help produce the belief that p , or evidential support for p , an intention to believe that p alone cannot get one to the belief that p . However, the obligation theorist is entitled to this account of the Uncontrollability Thesis, only insofar as she is entitled to the account of Transparency—which as we have seen, she is not.

The explanation given by the obligation normativist is one which explains a version of the Uncontrollability Thesis which holds as a matter of conceptual necessity. This is because the Transparency to truth considerations is a necessary feature of deliberation over what to believe.²³ As we saw in my description of Transparency earlier, Shah is careful to point out that Transparency does not rule out non-evidential factors influencing belief formation, but only that the relevance of such influences must go unrecognised by the subject in doxastic deliberation, it is not ‘that evidence has sole *causal* influence over belief’, but rather it has ‘sole *normative* or *rational* authority over belief’ (Shah 2003: 465). He suggests that unlike other accounts of deliberation over what to believe, his is able to acknowledge the ‘indisputable fact that beliefs can be influenced by non-evidential considerations’ and this is because his view ‘entails only that one is forced to apply belief’s standard of correctness in situations in which one exercises the concept of belief’ (Shah 2011: 96). But this is precisely what claim ii) in Noordhof’s impossible conjunction has occurring, the subject believes that p as a result of having the intention to believe that p , hence the subject recognises the influence on her belief formation of a non-evidential factor—it is precisely this which is ruled out by Transparency.

We find a similar line in Shah and Velleman (2005), in which they claim that:

[w]hen one deliberates whether to have an attitude conceived as a belief that p , one deliberates about an attitude to which one already applies the standard of being correct if and only if p is true, and so one is already committed to consider it with an eye exclusively to whether p . (Shah and Velleman 2005: 501)

So what is doing the work in explaining why the strong version of the Uncontrollability Thesis holds is an appeal to Transparency, and the part of Noordhof’s conjunction which is blocked is conjunct (ii). If, when deliberating over *whether to believe that p* , I apply the standard of correctness and am thereby committed to considering *whether to believe that p* exclusively with respect to

²³ In chapter one I distinguished between Transparency being a thesis about belief or Transparency being a thesis about deliberation over what to believe. I argued there that Transparency is most plausibly understood as a phenomenon characterizing the latter (§1.3.1).

whether p is true, I cannot believe that *p* merely as a result of having the intention to believe that *p*, because that is outside the bounds of those things I can consider as relevant to believing *p* in doxastic deliberation. However, the normativist cannot explain Transparency, and so she had better not appeal to Transparency in her explanation of the Uncontrollability Thesis. Rather, all she should say is something like the following: when I deliberate as to whether to believe that *p*, my reasoning is aimed at issuing (or not) in a belief that *p* in accordance with the norm of truth. To conceive of the resulting attitude as a belief, is to conceive of it as an attitude governed by a standard of correctness that a belief is correct if and only if it is true (Shah and Velleman 2005: 519). Given that the truth norm is ‘already applied to the envisioned attitude as part of the concept of belief’,

it must be the norm in accordance with which deliberation aims to issue in the attitude. Aiming to arrive at an attitude in accordance with a competing norm would be incompatible with subsuming it under the norm of truth in conceiving of it as a belief. (Shah and Velleman 2005: 519)

We see that the quoted material here does not include an appeal to Transparency, and yet can explain the Uncontrollability Thesis. The obligation normativist then can explain the Uncontrollability Thesis by appeal to a norm of belief, *even though* she cannot explain Transparency by this appeal.

The permissibility normativist can also give an account of the Uncontrollability Thesis, *even though* she cannot account for Transparency. The reason the permissibility theorist is entitled to the explanation of the Uncontrollability Thesis is because the norm of belief tells us that a subject may believe that *p* if and only if *p* is true. Considerations not relevant to determining the truth of *p* then cannot play a role in a subject’s discovering what she is permitted to believe. A subject may believe that *p* if and only if *p* is true, and so, given that we do not have control over whether a given proposition is true, we are not permitted to believe at will.

I have shown then that Transparency is not explanatorily prior to the Uncontrollability Thesis, unlike normativists have thought. It is not legitimate to appeal to an explanation of Transparency in order to explain the Uncontrollability Thesis, because no explanation of Transparency by appeal to a norm governing belief is successful. However, a norm governing belief *can* explain the Uncontrollability Thesis.

Let us look finally to the third feature of belief: Epistemic Normativity. I claimed that what required an explanation from belief theorists were the following two claims:

- (EN1) Beliefs have a standard of correctness.
- (EN2) Beliefs are governed by categorical epistemic norms.

I have said that my preferred account explains these features by appeal to biological normativity, such that there is no *sui generis* epistemic normativity which requires an explanation. The normative account will take claims (EN1) and (EN2) in their stronger versions, as involving irreducible epistemic normativity.

Of the three features of belief which indicate a link between it and truth, the explanation of Epistemic Normativity comes easiest to the normative account because according to it ‘the normative properties of belief are constitutive of belief, and are thus explained by the very nature of belief’ (Chan 2013: 8). The truth of claims (EN1) and (EN2) is going to follow from the very concept of belief. We learned earlier that it is part of this concept that a belief is correct if and only if it is true (claim (EN1)), and that the norm governing belief ‘is a constitutive and conceptual norm, which a subject needs to possess to be able to use the concept of belief’ (Engel 2013: 203), which gets us somewhat towards claim (EN2).

Other epistemic norms will not immediately follow from belief’s having a standard of correctness (it is conceptually coherent that beliefs might have this standard without there being other norms). The norm of truth alone does not exhaust the various features of Epistemic Normativity. Consider a case in which a belief is true, but something has still gone wrong, the subject comes to a belief that p where p is true accidentally (take any Gettier case, for example). Belief’s standard of correctness, captured by the truth norm, will not distinguish between the good case and a case like this because other epistemic correctness conditions do not follow from belief’s standard of correctness. Epistemic norms other than ones concerned with belief’s standard of correctness come into play here. To take an example from Ernest Sosa (2011), an archer can hit his target, and so his shot is successful if and only if it is correct. But ‘the shot can be more or less competent, apt or adroit. When we evaluate it, we do not attend only to the successful hitting—which could be just lucky—but also to the way it is performed, which can be good or bad’ (Engel 2013: 200). We can treat belief’s standard of correctness merely descriptively, without invoking any norms. Descriptive properties indicate what condition a certain kind has to satisfy in order to meet the given standard. Engel claims that the normative concept of correctness is distinct from this descriptive one. It concerns the way, or the operation which, an agent has to perform in order to meet the descriptive condition’ (Engel 2013: 200). The norm of truth, whether cast in terms of obligation or permission will not distinguish good cases from, say, Gettier cases, in both cases the subject does what she ought to. Other norms are required to draw distinctions between an agent’s having met the descriptive condition well, and an agent’s having met it badly. This is claim (EN2) of Epistemic

Normativity.

It does not matter what the strength of the norm governing belief is when it comes to giving an account of Epistemic Normativity. Either way, belief is constitutively (or conceptually) normative, and claims (EN1) and (EN2) are explained by appeal to this.

I will now move to look at an objection to the normative account which brings out the inappropriateness of trying to explain belief's connection to truth by appeal to a norm. *****

3.4 The No Guidance Objection

The No Guidance Objection claims that the norm of belief can offer no guidance to the believer. Given that norms are supposed to guide the activities they govern, a failure to do so speaks against there being a norm of belief. Here I will argue that the best response is given by an appeal to judgement, and that this response is available to both the obligation theorist and the permissibility theorist. However, the response reveals that an explanation of the connection between belief and truth cannot be found by looking to norms, and that, as with the teleological account, when we think in terms of triggering and structuring causes, the normative account is problematic.

Gluer and Wikforss raise this objection. They claim that a subject is told what to do by a guiding rule which operates under certain circumstances, for example 'Do X when in C'. All that is important for guidance is that the subject takes C to obtain, it does not matter if it is not transparent to the subject whether C obtains (Gluer and Wikforss 2009: 44). So for example, the rule 'carry an umbrella when it is raining', can guide my action, even if it is not transparent to me that it is raining, that is, even if I could be wrong about it raining. So long as I think that it is, the rule can offer guidance. There is a problem though when we look to the truth norm, with this feature of norms in mind.

Gluer and Wikforss focus on Wedgwood's formulation of the truth norm (outlined in 3.2.2), as they raise the same issues regarding Shah's formulation as those raised by Bykvist and Hattiangadi, namely, the Too Demanding Objection. They discuss then, the following formulation:

[(6)] If one considers whether p , then one ought to believe p iff p is true. (Gluer and Wikforss 2009: 43)

For this norm then, what takes the places of 'X' and 'C' in the 'Do X when in C' rule is 'believe p ' and 'if one considers whether p , and p is true' respectively. So what one should do is believe p

when in circumstances such that one considers whether p and p is true. The problem, though, is that in order for a subject to follow (6), she has to have a belief about whether p . If she considers whether p and forms the belief that p is true, then (6) tells her that she ought to believe that p . This gives the result that (6) ‘tells her that if she believes that p , she ought to believe that p ’, and as Gluer and Wikforss claim, it is ‘rather obvious that no guidance can be had from this’ (Gluer and Wikforss 2009: 44). If, on the other hand, the subject considers whether p and comes to believe that $\sim p$, (6) tells her that she ought not to believe that p . Once again, this is not to offer any guidance. Gluer and Wikforss conclude that ‘there cannot be truth-norms (of whatever form) that serve to guide our belief formation’ (Gluer and Wikforss 2009: 44).

The objection goes through for the permissibility account equally. Recall that the norm supposed to be governing belief on this account is:

[(5)] For any S, p : S may (believe that p) if and only iff it is true that p . (Whiting 2010: 216)

So for this norm, what takes the place of ‘X’ and ‘C’ in the ‘Do X when in C’ rule is ‘believe p ’ and ‘it is true that p ’. But note again that the doing of X is only something which is permitted and not required. So what one is permitted to do is believe p when in circumstances such that p is true. The problem is that in order for a subject to follow (5), she has to have a belief about whether p . If she considers whether p and forms the belief that p is true, then (5) tells her that she is permitted to believe that p . So again we get the result that the norm tells a subject that if she believes p , she may believe that p . On the other hand, if a subject considers whether p and sees that $\sim p$, the norm tells her that she is not permitted to believe that p , that is, that she ought not believe that p . Again, it is ‘rather obvious that no guidance can be had from this’ (Gluer and Wikforss 2009: 44).

Shah and Velleman claim that this objection arises from their explanation of Transparency:

[r]easoning cannot aim at issuing in an acceptance of p if and only if that acceptance would be correct in virtue of p ’s being true, because pursuit of that aim would entail first ascertaining whether p is true; and ascertaining whether p is true would entail arriving at a belief with respect to p , an intermediate step in deliberating whether to believe it. And believing that p cannot be an intermediate step in deliberating whether to believe that p . (Shah and Velleman 2005: 519–20)

They construe the No Guidance Objection as claiming that doxastic deliberation concludes with the following practical syllogism:

I will believe that p if and only if p is true.

P is true. (Shah and Velleman 2005: 520)

The problem is that the minor premise of the practical syllogism, ‘ P is true’, would be embodied in a judgement, a judgement which ‘would be equivalent to the one with which we imagine doxastic deliberation as concluding’ and so doxastic deliberation looks to ‘require its own conclusion as a premise’ (Shah and Velleman 2005: 520).

3.4.1 The Norm of Belief as an Idealization

Engel responds to the No Guidance Objection by biting the bullet: the truth norm governing belief ‘is indeed a conceptual norm, which does not give us any prescriptive—or even permissive—guidance’ (Engel 2013: 208). Constitutive norms are such that they do not prescribe, they are rather instances of an ‘ideal of reason’—the norm of truth tells the subject what she ought to believe (the truth), and is identified by Engel as falling into the category of ‘*ought-to-be*’ rather than ‘*ought-to-do*’ (Engel 2013: 208). The norm only informs us that we ought to believe truly, without suggesting how this might be done. Engel then rejects Gluer and Wikforss’s claim that norms take a ‘Do X when in C’ form. Rather we should say that

an ideal is meant to describe an abstract situation which holds only ‘in principle’, or a kind of conduct which only certain imaginary beings endowed with powers which are distinct from ours could follow (logical saints, believing all the consequences of their beliefs, perfectly rational agents). The status of the truth norm for belief, [...] is of this sort: it tells us what believing requires, but neither what kind of beliefs one must have before applying the norm nor what kind of belief one must have once one has applied it. It is blind to the actual psychology of the agents. In this sense, *it need not explain or guide our belief formation*. (Engel 2013: 209, my emphasis)

The interpretation of the truth norm given by Engel as an idealization means that the norm ‘does not entail that you have the belief in question, nor that you know whether it is available to you in any psychological sense’ (Engel 2013: 209). Engel discusses the norm governing belief as an ‘one ought to believe only what is true’ (Engel 2013: 209), which is equivalent to Boghossian’s proposed norm (§3.2.1) Engel does say that the norm cannot even offer permissive guidance, and so presumably he takes his remarks to apply to a merely permissive norm. Engel’s response should worry the normativist, particularly as he claims that the norm governing belief ‘need not explain or guide our belief formation’ (Engel 2013: 209). In this chapter I have been discussing the normative account in terms of how well it can account for the link between belief and truth, in particular, how well it can explain the three features of

belief outlined in chapter one. We saw that Shah (2003) and Shah and Velleman (2005) took themselves to be explaining the presence of Transparency (though this explanation does not work), and the explanation for why we cannot believe at will piggy-backed on that (though I suggested that an explanation could and should be given without an appeal to Transparency). If we understand the three features needing an explanation as being features of our belief *formation*, Engel's capitulation to the No Guidance Objection rules out those features being explained by a truth norm, whether couched in terms of obligation or permissibility. The normativist should look elsewhere for a response to the No Guidance Objection, in particular, she should seek to explain how the norm governing belief provides guidance to believers. Without such guidance, it is difficult to see how an appeal to the norm governing belief can explain the three features of belief and its formation outlined in chapter one.

However, this is not to say that Engel's response does not work, but rather, that he concedes too much to the opponent. He might want to concede that the norm cannot offer *direct* guidance, but he could still say something like the following:

[the] norm cannot provide direct guidance, but it may well issue in derived prescriptions about effective means to the end of satisfying it. Thus someone who is concerned to satisfy [the truth norm] by avoiding error may have reason, with respect to specific subject matters, to conform to such prescriptions as: 'believe *p* if and only if you have considered relevant alternatives', or 'believe *p* if and only if your informants are experts', and so on. (Papineau 2013: 66)

So the response to the No Guidance Objection could be to say that the norm does not provide direct guidance of the form 'Do X when in C', as Gluer and Wikforss demand, but it can suggest ways one might go about satisfying the norm. This does not run the normativist into the objection 'since it is perfectly possible to ascertain whether one has considered relevant alternatives, or has consulted experts, without yet having settled the question of whether *p*' (Papineau 2013: 66)

This is fine for answering the No Guidance Objection, but we might then ask how this move would allow the normativist to do the explanatory work in accounting for the three features of belief outlined in chapter one. Her position now is that the norm of belief does not directly guide, but may issue in prescriptions of ways to satisfy it. If this is right, then the normativist loses her (unsuccessful) explanation of Transparency. We saw earlier that the explanation of Transparency was given by the obligation theorist's appeal to dispositions activated when an agent frames her deliberation with the question *whether to believe that p*. This is because in exercising the concept of belief, one accepts that 'in some sense one *ought* to believe

that p only if p is true' (Shah 2003: 449). But if the norm of belief does not guide directly, but only issues in prescriptions regarding how best to believe that p if and only if p is true, why does the question *whether to believe that p* collapse into the question *whether p is true* in doxastic deliberation? If the norm of belief issues in a prescription such as *believe in line with the available evidence*, then this does not explain why we inescapably move from having lots of evidence pointing to p to believing that p in doxastic deliberation. Sure, if p is supported by the evidence, then I might be guided by the above prescription to believe that p . But if my following the evidence prescription is what explains my belief formation, then I ought to be able to not comply with the prescription. If we cannot disobey the prescription to believe that p when there is lots of evidence for p , we might think that this speaks against there being a genuine prescription to this effect (see Buleandra 2009). We cannot help but obey these kinds of prescriptions. If I recognise that p is well supported by evidence in doxastic deliberation, I cannot help but move to a belief that p . Now you might say that I can ignore the prescription insofar as I can withhold belief—I see that p is well supported by evidence, and withhold belief anyway. But recall the discussion of confidence thresholds earlier (§2.3.1.2)—our being able to do this does not suggest that we are not governed by Transparency, but rather that more evidence is required at sometimes rather than others. Thus, answering the No Guidance Objection by appeal to derived prescriptions just pushes the problem back. Now the normativist must say explain how these prescriptions can guide, and she will not be able to do this if those prescriptions are also going to be able to explain Transparency, because Transparency indicates that the prescriptions cannot be ignored, and we are motivated inescapably to believe in accordance with them. And if this is right, then there is a question over whether they are *prescriptions* after all.

3.4.2 Judgement that p

Shah and Velleman take an alternative strategy in their pre-emption of the No Guidance Objection, and opt for the norm of belief *directly* guiding believers. They argue that we do not deliberate about *whether to believe that p* by considering *whether p is true*, but rather by considering *whether p* . It is by considering *this* question that we aim at arriving (or not) at a belief that p in accordance with the norm of truth (Shah and Velleman 2005: 520). If one judges that p , that judgement

will induce an acceptance of p in a way that, to the best of one's ability, accords with the standard of correctness implicit in the concept of belief and so it will, to the best of one's ability, accomplish the aim characteristic of doxastic deliberation. (Shah and Velleman 2005: 520)

Considering the question *whether p* then is the way to accomplish the aim of believing in accordance with the norm of truth. In order to consider this question as a means to answering the question *whether to believe that p*, one need regard answering it as a means conducive to the aim of doxastic deliberation, and so one ‘must regard it as likely to issue in a judgement that *p* if and only *p* is true’ (Shah and Velleman 2005: 520). The response then is to claim that when one is focused upon *whether p*, one is indirectly committed to the norm of truth. When considering *whether p*, and following truth conducive methods, one manifests one’s responsiveness to the norm of truth governing belief.

I focus on *whether p* when I enter into doxastic deliberation, because I am committed to the norm of truth, which tells me that I ‘ought to believe that *p* only if *p* is true’ (Shah 2003: 449). This is why I focus on *whether p* when I deliberate about *whether to believe that p*. So the norm of truth does guide me, not by suggesting that I first find out whether *p*, and if *p* is true then I can go ahead and form the belief that *p*. Rather, it is on *whether p* that I focus when I deliberate about *whether to believe that p*, and so I indicate that I am committed to the norm of truth, because immediately what I do when I deliberate about *whether to believe that p*, is focus on *whether p*. The fact that I do this displays my commitment to the norm.

What we learn from the No Guidance Objection is that doxastic deliberation is not directly aimed at the truth—in order to form a correct belief about whether *p*, ‘one cannot aim in the first instance at accepting *p* if and only if it is true; one must aim at following some truth-conducive methods that will lead to its acceptance’ (Shah and Velleman 2005: 520). Such methods are those conducive to answering the question *whether p*. The No Guidance Objection then, can be answered by the normativist by claiming that in considering *whether p* in doxastic deliberation, one manifests one’s commitment and responsiveness to the norm of truth governing belief.

3.4.3 Permissiveness and Guidance

We have seen that because of the Too Demanding Objection, the normativist must move to a permissive norm, and so it had better be the case that the permissive norm can also answer the No Guidance Objection. The permissibility account is entitled to the response offered by Shah and Velleman above. When I enter into doxastic deliberation, I display my commitment to the norm that I ‘may (believe that *p*) if and only if it is true that *p*’ (Whiting 2010: 216) by immediately attending to *whether p*. So it is not that I look to *whether p*, ascertain *whether p*, and then apply the rule that I ‘may (believe that *p*) if and only if it is true that *p*’. Rather, the very fact that I focus immediately on *whether p*, displays my commitment to the norm that I ‘may (believe

that p) if and only if it is true that p . If I have no disposition to move to the belief that p except by apprehension of the reasons for believing that p , I display my commitment to the norm of truth governing belief. The No Guidance Objection then can be answered by both the obligation and the permissibility theorist.

However, though this line answers the No Guidance Objection, the way it does so is indicative of a different problem. This response to the No Guidance Objection reveals that an appeal to truth to explain a triggering event (the disposition to move immediately from *whether to believe that p* to *whether p is true*), makes no sense. We should not locate the role of truth at an intentional level—with the subject exercising the concept of belief. Rather, we should look to the sub-intentional level which reveals a pre-commitment to truth, secured by neural structures, which can be explained by an appeal to biology.

I indicate my commitment to the truth norm when I move immediately to *whether p is true* when I consider *whether to believe that p* . But now we have the same problem as faced by the teleological account. If it is my pre-commitment to the norm of truth which guides my belief formation, which makes it the case that I attend to *whether p is true* when I enter into doxastic deliberation, how does this translate into the language of triggering and structuring causes? Is my commitment to the norm a triggering cause of why I move to *whether p* when thinking about *whether to believe that p* ? Is my commitment to a norm what triggers the disposition to allow certain considerations to settle whether or not I should have a belief that p ? But then what caused my commitment to the norm? The normative account is silent on this. Perhaps the normative account is compatible with a biological approach: the commitment to the truth norm is the triggering cause of our immediately attending to *whether p* in doxastic deliberation, and what secures why we are committed to this norm, is biology. However, if the commitment to the norm of truth is the triggering cause of why we immediately attend to *whether p* , presumably we can, again, give a lower level causal description of what causes this attending, a causal description which appeals to the neurological structures which secure this disposition, and then talk of commitment to the truth norm is explanatorily redundant. On the other hand, if we accept that some causal facts can explain why we attend to *whether p is true* when considering the question *whether to believe that p* , then we only need to appeal to biology for the structuring cause of these causal facts, and there is no explanatory role for a norm to play. If we try to make a place for the norm in the causal picture, it does not fit well as the structuring cause for why we move to *whether p is true*, because it does not look plausible to claim that our commitment to the norm secures the causal facts which moves us from *whether to believe that p* to *whether p is true* (how would recognition of a norm cause certain neurological structures to be instantiated?). So, as

with the teleological account, when we are thinking in terms of triggering and structuring causes, and locating the role of truth in the triggering cause, the notion of a norm as explanatory does not sit comfortably in this story.

My preferred account has truth coming into an explanation of the three features of belief as a structuring cause, we have seen from both the teleological account and the normative account that an appeal to truth as a triggering cause does not work. An appeal to biological function can instead play the explanatory role.

3.5 Conclusion

In this chapter I argued that because of the Too Demanding Objection, the normative account of belief needs to be couched in terms of permissibility and not obligation. I discussed how this account might explain the three features of belief which indicate a link between it and truth. I argued that Transparency could not be explained by appeal to a norm, but the Uncontrollability Thesis could be, which demonstrates that Transparency is not explanatorily prior to the Uncontrollability Thesis. I then argued that though the No Guidance Objection could be answered, the answer to it showed that the normative account does not work when we think about triggering and structuring causes. Accounting for the link between belief and truth by appeal to a norm then, is inappropriate. Therefore, the normative account cannot give an explanation of the connection between belief and truth and so we ought to reject it.

THE ETIOLOGICAL ACCOUNT OF BIOLOGICAL FUNCTION

In this chapter I will defend the etiological account of biological function. After outlining the account in its current form, I will pose Paul Sheldon Davies's Impossibility of Malfunction Objection as a challenge to the etiological approach. If right, Davies's objection shows that the etiological approach cannot ground a notion of biological normativity and thus cannot explain malfunction. This means that normativity cannot be based in natural selection in the way that I claim it is in my account of belief. In response I will develop the etiological account such that it can answer Davies's objection. I will also argue that my proposal provides a response to Bence Nanay's Circularity Objection.

It is beyond the scope of this thesis to argue that the etiological account is the only viable account of biological function, or even the best one. Rather, I will argue here that it is *an* option, insofar as the objections raised to it can be answered. This will allow me to adopt the account when formulating my account of belief in the next chapter. I will explain the existence of the three features of belief identified in chapter one, by appeal to the biological functions of the mechanisms for their production. Given this, only a historical account of biological function can do this work because historical accounts take as their explanandum why a functionally characterized trait exists, or in the case of belief, why a trait has the nature that it does. *****

4.1 Historical Accounts and the Connection Between Belief and Truth

Non-historical accounts of biological function do not claim that function ascriptions explain the presence of a trait. The Systemic approach claims that 'the function of an item is the role of that item in bringing about an activity or capacity of a complex system of which that item is a part' (Wouters 2005: 135). Robert Cummins, a proponent of this approach, claims that what is explained by function statements are the 'biologically significant capacities of an entire organism' (Cummins 1975: 760), and rejects the claim that it is 'the presence of [an] item [...] that is functionally characterized' (Cummins 1975: 741). Life chances approaches claim that '[i]n describing the function of some biological character, we describe some presently existing item by reference to some future event or state of affairs' (Bigelow and Pargetter 1987: 181). Finally goal contribution approaches claim that functions are 'causal contributions to the maintenance of a goal state', with the criteria for a system's having a goal being that it 'is disposed to vary its behavior in the manner required to achieve or maintain the goal' (Wouters 2005: 138). This is

not to say that proponents of these accounts would disagree with my function ascriptions (they may well say that one of the biological functions of our mechanisms for belief-production is the production of true beliefs, for example), but their having this function will not explain the features of belief which I want to explain. Claiming that the function of our mechanisms for belief-production is the production of true beliefs, will rather, on these accounts, explain a biologically significant capacity (systemic approach), describe the mechanisms as ones which will enhance the life chances of its bearers (life chances approach), or describe the production of truth beliefs as being a contribution to the maintenance of some goal (goal contribution approach). An appeal to biological function as conceived under any of these accounts, will not explain *why* beliefs have the features that they do.

We might usefully understand the difference between a function ascription on historical and non-historical accounts as offering structuring and triggering causal stories respectively. When a historical theorist ascribes the biological function of producing true beliefs to our mechanisms for belief-production, that ascription offers a structuring explanation of why certain causal facts about our belief formation hold. When a non-historical theorist offers this same function ascription she offers a triggering explanation of some phenomenon. There are two different explanatory tasks. The first is to explain why certain facts about belief hold, those facts which constitute the basis of Transparency, the Uncontrollability Thesis, and Epistemic Normativity. The second is to explain why belief is characterized by these features at all. Again, the first explanation is a triggering explanation, the second is a structuring one. Non-historical accounts will answer the triggering question, my preferred historical account will answer both by giving a structuring explanation of the triggering causes. The only way in which truth in terms of belief is explanatory is related to structural explanation. A non-historical approach which offers a triggering cause when it ascribes a function does not explain the role of truth in belief formation, but only how belief's connection to truth is realised. On a historical account though, I can explain the presence of mechanisms related to truth by appeal to their having a biological function, where such an ascription is taken to explain why those mechanisms exist.

4.2 The Problem of Function

It is important to get clear on what the proponents of function accounts are up to in offering them. If we correctly construct a theory of what biological functions are and on what grounds they are to be ascribed, what work can that theory do?

There is fairly wide consensus on the thought that function statements are inherently puzzling, owing to their being both explanatory and normative, with respect to this Valerie Hardcastle claims that:

[f]unction is a strange concept: normative, on the one hand, and completely naturalistic, on the other. The word both describes what a thing is supposed to do and tells us something about the way the world works. (Hardcastle 2002: 144)

Carolyn Price also gives special attention to the explanatory value and normativity of function statements, claiming that they give rise to a puzzle. She thinks that function claims are explanatory—they explain why a trait exists now²⁴—but that they are also normative, insofar as they are claims about what traits are *supposed to do*. Price thinks that though these features are not puzzling in themselves, the combination of them gives rise to a difficulty: the explanatory value we get from function claims is grounded in what an item did in the past, which explains why it is here now, such explanatory force suggests that function statements are backward-looking. On the other hand, the normative aspect of function claims supports the view that function statements are forward-looking, they tell us what it is an item is supposed to do now, or in the future (Price 2001: 11–12). In other words, these features suggest the following question: ‘How can knowing what a device is supposed to do help explain the fact that it is there doing it?’ (Price 1995: 144).

Many other philosophers have claimed that normativity is a key feature of function. Arno Wouters suggests that ‘the main task of a theory of function is to explain how this norm arises in biological contexts’ (Wouters 2005: 124). Peter McLaughlin also claims a special place for the normativity of function statements, stating that ‘[t]he fault line running through [the function] debate seems to follow the question of norm and value’ (McLaughlin 2001: 4). To claim that something has a function is to ‘view it as a means to an end, as instrumental or useful for something that itself is valued or somehow normatively distinguished’ (McLaughlin 2001: 4). Finally, William Wimsatt suggests that the ‘attribution of function to an entity implies that the entity promotes the meeting of some standard of evaluation’ (Wimsatt 1972: 8).

If it is right that normativity is a key feature of functional items or statements, the etiological account does rather well here, as it can accommodate this feature. Furthermore, in

²⁴ There are philosophers who object to the idea that function claims are explanatory in this sense. Non-historical theorists do not think that function ascriptions explain the presence of an item, but something else, as I discussed in §4.1.

adopting the etiological account in giving a biological account of belief, the normativity of belief is accounted for (insofar as there is a standard laid down by the past performance of our mechanisms for belief-production, from which the products of current mechanisms can deviate). In adopting the etiological account of biological function, we can account for the normativity of belief, without running into the problems the normative account faced, as discussed in chapter three.

There are more issues which arise for biological function which theories thereof ought to account for, to name just two: an account of biological function ought to offer the tools needed for biologists to ‘identify organs between individuals and between species’ (Wouters 2005: 127). Also, an account of biological function ought to be able to distinguish an item’s having a function from an item’s performing a function (Wouters 2005: 128). This distinction could also be understood as the distinction between functional performance and accidental performance. For Larry Wright, making sense of this latter distinction, is ‘perhaps the primary aim’ (Wright 1973: 142) of the account of function he develops.

A theory of function should also be able to distinguish between activities of a trait which are functions of that trait, and activities which are side effects or accidentally useful. It should also allow us to ascribe functions to items which are unable to perform them, such as malformed items (Wouters 2005: 133). Having now a better idea of the explanatory burdens facing the function theorist, I will move on to describe, in broad terms, two types of etiological account.

4.3 Etiological Accounts²⁵

Broadly speaking, there are two types of etiological account, which I will refer to as the *strong etiological account* and, following David Buller (1998), the *weak etiological account*. Both accounts ground function ascriptions in the causal-historical relationship held between present day tokens and ancestral tokens of a given trait type. The difference between them is that the strong

25 Exegesis of the etiological account typically starts with Wright’s account developed in his 1973 paper ‘Functions’, and his 1976 book *Teleological Explanations*. In these works, we find the claim that to say that the function of an item X is Z is to say two things. Firstly, it is to say that ‘X is there because it does Z’ and secondly that Z is a consequence (or result) of X’s being there’ (Wright 1973: 161, 1976: 81). Wright argues that this formulation captures the explanatory component of function statements, when we ascribe to an item a function, we are capturing ‘the particular consequence’ of that item’s being where it is, and that consequence explains why that item is there (Wright 1976: 78). Wright offers the example of the second hand on his watch claiming that though it may keep the watch face clean by sweeping away dust, that is not the reason the second hand is there, and so that is not its function. The reason it is there is because it assists in reading seconds, so that is its function (Wright 1976: 78). There is a little contention over whether Wright is correctly characterised as an etilogist (see Millikan 1989a: 176, n. 1, and a reply to this in Neander 1991b: 459, fn. 8).

account grounds its function claims in *selectional* history, which requires variation, the weak account does not. So on the strong account the function of an item is determined by that item's history, more specifically, its selectional history: 'the purpose of A is to do B if and only if A is now present because in the past some selection process selected items that do B' (Papineau 1993: 59). Karen Neander sees this as the 'central element' of the [strong]²⁶ etiological approach, claiming that 'the function of a trait is the effect for which that trait was selected' (Neander 1991b: 459).

In contrast, the weak etiological account claims that:

[a] current token of a trait *T* in an organism *O* has the function of producing an effect of type *E* just in case past tokens of *T* contributed to the fitness of *O*'s ancestors by producing *E*, and thereby causally contributed to the reproduction of *T*s in *O*'s lineage. (Buller 1998: 507)

On the weak etiological account variation is not a necessary condition on function possession, that is, a trait *T* can have the function to *F* providing '*T* contributed to the fitness of the ancestors of *T*'s current bearers *by* producing an effect *E* and that this, in turn, contributed to the reproduction of *T*s' (Buller 1998: 508). The weak etiological account is more liberal than the strong etiological account, which requires variation for function ascription. On the strong account, for a trait *T* to have the function *F*, 'an item must not just contribute to survival or reproduction, it must contribute to reproduction better than some alternative item, that is, it must contribute to differential reproductive success or fitness [...] Not only the history of the function bearer is relevant to the ascription of biological proper functions but also the history of competing items that have fleetingly come and gone' (McLaughlin 2001: 106).

In this chapter I will not distinguish between the two accounts, insofar as I will defend them both. The differences between them are not important for the theoretical work I put the etiological account to. I think that the strong etiological account is right, but I need not commit to it—the weak will do for my purposes. The etiological account, in its most general form, accepts that function claims are both explanatory and normative, and seeks to give an account in light of these features. The objections I will consider have as their target any account of function which has it that a necessary and sufficient condition on function ascription to a trait, is that

26 I enclose 'strong' in square brackets to indicate that the distinction had not been made at this point—Neander did not claim to be a strong etiological theorist, but she qualifies as one on Buller's taxonomy.

trait's having an appropriate history. It is this which characterises the both etiological approaches, and upon which all etiological theorists agree.

The etiological account claims that an item's biological function 'is to do whatever items of that type did that caused them to be favoured by natural selection' (Neander 1991a: 174) (strong) or what contributed to some component of organismic fitness (weak). Claims about biological function are made at the level of types, not tokens, and so 'the forward-reference to a trait's²⁷ function, to what the trait is supposed to do, serves as an implicit reference to past selection of that type of trait for that type of effect' (Neander 1991b: 461). As well as the forward-reference to a trait's function, there is a backward-reference to a 'causally explanatory selection process, during which those items or traits were selected for those effects which are their functions' (Neander 1991b: 467). Again, a weak theorist can recast these claims in terms of the trait contributing to some component of fitness without requiring that it was selected over historical traits no longer present.

Characterized in this way, the etiological account is too liberal. If descent from a successful lineage is claimed to be sufficient for function attribution, the etilogist finds herself in the uncomfortable position of attributing functions to vestigial traits. In light of this, the etiological account needs to adopt a modern history clause which will rule out vestigial traits being functionally characterized items. Griffiths suggests that the conditions under which the etiological account attributes function, must apply to a trait during the most recent 'evolutionary significant period'. An evolutionary significant time period for a given trait T is defined as:

a period such that, given the mutation rate at the loci controlling T and the population size, we would expect sufficient variants for T to have occurred to allow significant regressive evolution if the trait was making no contribution to fitness. (Griffiths 1993: 417)

For Griffiths, if a trait has a function in the present day, selection for it needs to have occurred in the last evolutionarily significant time period (Griffiths 1993: 417). Peter Godfrey-Smith, as

²⁷ There is a question about whether 'trait' here refers to a trait in an individual or a trait in a species. Buller claims that the function literature has not been clear on this issue, with two types of explanation being claimed for the etiological account. On the one side, one could take function ascriptions to be explaining why a trait token *in its bearer* is present, or, on the other, one could take function ascriptions to explain why a trait type *in a species* is present. Buller points out that Neander (1991b) seems to subscribe to the former, whereas Millikan (1989a), Paul Griffiths (1993), and Colin Allen and Mark Bekoff (1995) subscribe to the latter (Buller 1998: 522). I do not need to distinguish between the views here. I am appealing to biological function to explain features of our mechanisms of belief formation. Function ascriptions can be taken to apply to a particular believer (but will also apply to all believers) or to apply to all human believers.

part of his ‘modern history’ approach to function, makes a similar move, and refers to the ‘recent maintenance of [a] trait under natural selection’ (Godfrey-Smith 1994: 344). His account is an etiological one; when we ascribe a function to a trait, we are making a claim about the past, ‘but the relevant past is the *recent past*,^[28] modern history rather than ancient’ (Godfrey-Smith 1994: 344, my emphasis).

We can now see how the etiological approach can do the explanatory work required of an account of function outlined in §4.2. There I referred to Wouters’s claim that accounts of biological function ought to have the resources to ‘identify organs between individuals and between species’ (Wouters 2005: 127). According to Millikan, biological kinds are ‘defined by reference to historical relations among the members’ (Millikan 1999: 54). What is really important to draw out here, is that membership to a biological kind, is not determined by properties of a given item, but by its history.²⁹ Of course, members of the same biological kind will share many properties, and this too is grounded in the history of those items. The reason members of the same biological kind are like each other in many respects, is the historical relations they bear to one another, where each item ‘exhibits the properties of the kind because other members of that same historical kind exhibit them’ (Millikan 1999: 54).

The etiological account can distinguish activities that are functions from activities that are side effects by appeal to selectional history (strong), or by appeal to which activities contributed to some component of organismic fitness (weak). It is only those activities which were selected for, or which contributed to fitness, which count as functions. Equally, the effects which are functions are those which were selected for (strong), or those which contributed to fitness (weak).

4.4 The Impossibility of Malfunction Objection

In this section I will take Davies’s objection to pose a serious problem for the etiological approach to biological function. My account of belief appeals to the biological functions of our mechanisms for belief-production to explain the three features of belief outlined in chapter one (§1.3). Importantly, I think that epistemic normativity is reducible to biological normativity, such

28 Godfrey Smith adds: ‘[s]ome might wonder how recent selective episodes relevant to functional status have to be. The answer is not in terms of a fixed time – a week, or a thousand years. Relevance fades. Episodes of selection become increasingly irrelevant to an assignment of functions at some time, the further away we get’ (Godfrey-Smith 1994: 356).

29 Though in my conditions on membership to a functional kind, to be introduced in §4.5, I introduce the requirement of one of a set of structural properties, though, as I will explain, I do not think this move is one which would be resisted by etiological theorists.

that the normativity of belief can be given by a naturalistic account. If Davies is right however, there is no biological normativity, and so it certainly cannot be something to which something else (epistemic normativity) can be reduced. Here I outline Davies's objection, before developing the etiological account in the next section in order to reply to the objection.

Davies has, in several places, launched a systematic attack on the etiological account, calling it a 'failed theory', which 'does not have the virtues that advocates claim for it' (Davies 2001: 32), and has 'rather little [...] worth preserving' (Davies 2001: 39). Davies wants to 'reject the claim that some natural objects possess intrinsic norms of performance' (Davies 2001: 103), arguing that '[a]s naturalists, we should eschew the attribution of norms of performance to natural, nonengineered traits' (Davies 2001: 208). The Impossibility of Malfunction Objection first appeared in his paper 'Malfunctions' in 2000, in which he did not propose an alternative view, rather, his aim was merely to motivate the need for an alternative, to 'clear the ground' (Davies 2000: 20). Later though, he restated the objection in his book *Norms of Nature*, and here provided an alternative account, according to which:

functions are nothing more than systemic capacities that contribute to the exercise of higher-level capacities we wish to understand and control. Functions [...] are effects that play a role in the workings of hierarchical systems. (Davies 2001: 5–6)

I will not consider the merits of Davies's account here, except to say that once again, appealing to biological function understood in this way cannot help explain the existence of our belief-producing mechanisms, because Davies rejects the thought that function claims explain why a trait exists. I will try to undermine the motivation for Davies's account by defending the etiological account from the charge that it cannot ground a notion of biological normativity and so it cannot accommodate malfunction.

As we have seen (§4.2), much of the biological function debate has been concerned with how to account for the supposed normativity of function. Often what the function theorist's burden comes down to is the requirement to accommodate *malfunction*. How can we understand the thought that there is something that biological items are *supposed* to do, that biological items can fail to do what they are *supposed* to do and that they can therefore *malfunction*?

Malfunctions are ‘ubiquitous’ (Davies 2000: 20) owing to the fact that ‘[b]iological functions are not, in general, reliable functions. They quite standardly go awry’³⁰ (Millikan 1995a: 175). The burden of accounting for malfunction is especially important for the etiological account, making Davies’s charge a serious one. This is because one of the supposed strengths of the etiological account is its being able to accommodate biological normativity and malfunction. It is a ‘persistent boast’ (Davies 2000: 19) of the etiological account, or indeed, ‘perhaps the grandest’ of its claims, that it can provide ‘a clear and compelling naturalistic explanation for the occurrence of malfunctions’ (Davies 2001: 190). Furthermore, it is claimed by proponents that *only* a historical account can do this work, because ‘nonhistorical analyses of “function” fail to deal adequately with items that are not capable of performing their functions’ (Millikan 1989b: 288). This is because such accounts ‘run afoul exactly when they confront the most central issue of all, namely, the problem of what failure of purpose and defectiveness are’ (Millikan 1989c: 299).

Proponents of the etiological account have been very clear that their preferred account can do this work. They claim that a historical connection to prior tokens of a biological trait is sufficient for a trait’s having a function and sets the norms on proper functional performance (Millikan 1989c: 299), that is, ‘being preceded by the right kind of history is sufficient’³¹ to set the norms that determine purposiveness’ (Millikan 1989c: 289). Biological traits then have functions ‘whether or not [they are themselves] capable of serving any of these functions’ (Millikan 1995d: 56)—a trait can make no causal contribution to fitness, but it nevertheless still has something that it is supposed to do (Neander 1991b: 459, fn. 7). Malfunctioning traits are those which ‘have functions that they are supposed to perform, but which they lack the disposition to

30 There is an important distinction between an item *failing to perform* its function, and an item *malfunctioning*. In later work, Millikan is more doubtful about whether anything falls into this second category, and so she thinks that the etiological account need not concern itself with accounting for malfunctioning items—this work can be done by appeal to abNormal conditions. So we should not say that an item is malfunctioning, rather we should say that an item is failing to perform its function because conditions are abNormal for proper functional performance. Given this story, Millikan thinks that malfunction proper is actually very rare (if indeed it happens at all), and that an item failing to perform its function is the ‘usual case’ (personal correspondence). I will put this aside for the purposes of this chapter. Theorists differ with respect to the extent to which malfunction is possible, and its prevalence. For those who think it is possible, and that it is something which an account of biological function ought to accommodate, I offer a way the etiological account might do so. The point here is to show that the etiological theorist can—if she deems it necessary—account for malfunctioning items. I will also write as though failure to perform function is malfunction (in line with Neander (1991b)). I do not think this is right, but the wider the etiological theorist is able to cast the malfunction net, the better position she is in to respond to Davies.

31 We will see (§4.5) that this needs some qualification. It is by qualifying this claim that malfunctions are possible on the etiological account.

perform' (Neander 1991b: 466, fn. 13). When a trait fails to perform its proper function, that is, when a trait fails to do what it is *supposed to do*, the etiological account claims that a trait has malfunctioned.

As we saw, Davies's larger aim is to reject the claim that there is an explanatory burden on the function theorist to explain the normativity of function. As such, being able to account for the supposed normativity is not a point in favour of an account. There might be a different kind of explanatory project with which the function theorist ought to engage, perhaps that of giving an account of the 'urge to attribute norms that violate our naturalistic commitments' (Davies 2001: 15). Davies thinks this is a project best undertaken by a study of the nature of such an urge, but that this is all there is to explain.

Davies argues that from a naturalistic perspective, it is difficult to see 'how or why we should think that the processes that result in trait preservation confer upon descendent tokens a norm of performance' (Davies 2001: 67). I will show that the etiological account can, *by its own lights*, accommodate biological normativity and malfunction. I will claim that Davies's argument teaches us that the etiological account needs to be formulated more precisely with regards to the details of the causal-historical relation between present day tokens and ancestral tokens of a functional type. Once the account does this in line with my proposal, it lay down a standard from which token traits can deviate and accommodate malfunction.

Davies's argument for the claim that historical malfunctions are impossible is, in summary:

[i]f functional types are defined in terms of historical success, then tokens that lack the defining property due to defect, and tokens that have lost the defining property due to disease or damage, are excluded from the functional category. Historically based malfunctions, in consequence, are impossible. (Davies 2000: 19)

The primary target of Davies's objection is the strong etiological account, the argument is applied to the weak etiological account only derivatively, as we will see. For now then, I will explicate the objection in terms of the strong etiological account. Davies draws from the following exposition of the strong etiological account (what he calls the 'Selected Effects' account): for organism O and trait T, in selective environment E, the function in E of T in O is to do F if and only if:

- (i) Past instances of T in O performed F in E,
- (ii) T was heritable,
- (iii) Past performances of F caused an increase in O's relative ability to satisfy demands of E (relative to other organisms in the population lacking T),

- (iv) This increase in O's ability to satisfy selective demands of E resulted in an increase in O's long-term relative rate of reproduction,
- (v) This increase in relative reproduction resulted in the persistence or proliferation of O and hence tokens of T. (Davies 2000: 22, Davies 2001: 19)

To prevent the ascription of functions to vestigial traits, Davies revises the characterization of the strong etiological account so that conditions on a trait's having a function appeal to 'genealogical descent from a *recently* successful lineage' (Davies 2000: 25, my emphasis). Davies then characterizes the strong etiological account thus:

[a] token of trait T has selected function F if and only if the token is descended from a lineage perpetuated by the recent selective efficacy with which ancestral members performed F. (Davies 2000: 25)

Conditions (i)–(iv) then, refer to *recently* successful ancestral tokens of T. Finally, Davies stresses that the 'conditions under which a token trait has a selected function definitely *do not* include possession of the capacity required to actually perform the functional task' (Davies 2000: 25). This is of course granted by, and indeed is one of the insights of, etiological accounts.

In setting up his objection, Davies distinguishes *generic trait types* from *selected functional types*. The first category is broader, including all traits of a given kind which are apt for selection. The latter is narrower, including only those variants of a given generic trait type which are selectively successful, which possess the property selected for. In the oft-cited case of Kettlewell's (1973) peppered moth³² for example, the targeted trait is wing colour, and this is the generic trait type. The successful variant of this type is a particular wing colour, and it is this which is the selected functional type.

With this distinction in mind, Davies now asks which of these types tokens of T cited in (v) are members of. This is hugely important: if tokens of T in (v) are members of the generic trait

³² Before 1850, near Manchester, tree trunks roosted on by peppered moths were lichen-covered. This meant that those moths with light coloured wings were protected through camouflage, from predation from birds, whilst moths with dark coloured wings were heavily preyed upon. This was until industrial soot destroyed the lichen, which resulted in a darkening of the tree trunks. By the end of the nineteenth century, the selective successes of the moths had been reversed. Now the dark coloured winged moths were protected from predation, and the light coloured winged moths were exposed to it. Whereas before 99% of the peppered moth population had light coloured wings, now 99% of the peppered moth population had dark coloured wings (Kettlewell 1973, cited in Davies 2000: 25).

type only, then the etiological story about malfunction goes through, because the generic trait type will include instances of the trait which are not selectively successful. So, tokens of T cited in (v) could refer to successful tokens, or unsuccessful (i.e., damaged or diseased) tokens. The latter, which could be understood as *malfunctioning* tokens, are nevertheless still part of the functional category, and so there is still something that they are supposed to do. This would be a good result for the etiologist.

However, Davies claims that because tokens of T in (i)–(iv) refer to those tokens which possessed the properties which were selected for and are thus members of both the generic trait type and the selected functional type, so too—on pain of equivocation—must tokens of T in (v) be members of both the generic trait type and the selected functional type. To avoid equivocation, tokens of T in (v) must refer not just to members of the generic trait type, but also to members of the selected functional type. This means that ‘membership in a selected functional category, per condition (v), requires possession of the property selected for’ (Davies 2000: 27). What follows is that a current incapacitated token, lacks the capacity which defines the selected functional type and thus, despite its being descended from a recently successful lineage, does not qualify as a member of the relevant functional type (Davies 2001: 203).

As a result of this, the etiological story needs to be revised thus:

[a] token of trait T has selected function F if and only if (1) the token is descended from a lineage perpetuated by the recent selective efficacy with which ancestral members performed F, and (2) the token possesses the property selected for in terms of which the relevant functional category is defined. (Davies 2000: 30)

On this formulation, the strong etiological account does not leave room for malfunction. This is because if ‘due to some sort of defect, disease, or damage the token of trait T lacks the property selected for, then it is merely a member of the generic type, not the functional type’ (Davies 2000: 30). If this is the case, norms of functional performance do not apply to such tokens. Davies concludes then that *selected malfunctions*, the only malfunctions available to the strong etiological account, are impossible.

The objection also works against the weak etiological account. Davies suggests that the eschewal of the appeal to selection ‘may give the appearance that [his] argument [...] does not apply to the theory of weak etiological functions’ (Davies 2000: 34). However, this would be to make a mistake as, he claims, given certain adjustments, the argument applies here also. The adjustment is to cast the argument not in terms of selective success as before, but more generally in terms of properties of historical success (Davies 2000: 34).

Even though for the weak etiological theorist variation is not necessary for functional kind membership, for any case, there either was variation among ancestral instances of the trait in question, or there was not variation, Davies argues that whichever is the case, his objection goes through. If there was variation, the etiological functional category gets defined by reference to the successful variants which contributed to ancestral fitness by performing F . So, as before, if a present day token does not possess the capacity to perform F due to defect, disease or damage, that token is not a member of the functional kind, and so it cannot malfunction. If there was not variation, the objection still goes through. Davies supposes that ‘not a single variant was sufficiently defective, diseased, or damaged to prevent it from contributing to fitness’ (Davies 2000: 34). Now the etiological functional category gets defined by reference to the tokens that contributed to ancestral fitness by doing F , in this case it just so happens to be all ancestral tokens. But again, present day tokens which lack the capacity to perform F , do not count as members of the functional type and so there is nothing that they are supposed to do. So weak etiological malfunctions, as well as selected malfunctions, are impossible (Davies 2000: 34). The etiological account then, in neither its strong or weak version, can account for malfunction. Without being able to do this, it is difficult to see in what sense it grounds a notion of biological normativity. Any standard laid down by past performance of a function can only apply to traits which are able to perform that function. Traits which might be candidates for deviating from that standard cannot do so without also failing to be a member of the functional kind, and thus failing to have this standard bestowed upon them. If the etiological account cannot ground a notion of biological normativity, the account I develop in chapters five and six, given that it adopts the etiological account, will not be able to reduce epistemic normativity to biological normativity.

4.5 Developing the Etiological Account

In its current form, the etiological account (both strong and weak) is vulnerable to Davies’s objection. My claim in this chapter is that the etiologist will be in a position to reply to Davies if she gets clearer on the causal-historical relation she cites which holds between present day tokens and ancestral tokens of a given trait type. Given this, I offer a more precise characterization of the causal-historical relation cited by etilogists. My proposal is the following: a token trait t is a member of the functional kind K and has F -ing as its selected

function if and only if it³³

- (a) Possesses one of a set S of intrinsic structural properties, $\{p^1, p^2, p^3, \dots\}$
- (b) Stands in a causal historical relationship to at least one token trait which
 - [i] possessed one of a set S of intrinsic structural properties, $\{p^1, p^2, p^3, \dots\}$
 - [ii] possessed one of a set S $\{p^1, p^2, p^3, \dots\}$ responsible for that token trait F -ing
 - [iii] was selected for F -ing.

Selection for requires variation, in line with the strong etiolist's claim that functions are attributed to traits on the basis of what they did which 'accounts for [their] presence in the population, as over against *historical* alternative traits no longer present' (Millikan 1995c: 40). For the weak etiological account, the formulation will only differ from that of the above by condition (b) [iii]. The weak etiological account does not require variation for function attribution, but only that a trait contributes—by F -ing—to some component of organismic fitness, and it is this requirement which constitutes clause [iii] in (b) for the weak etiological account.

If a token trait meets the conditions set out here, it is a member of a functional kind, and can be attributed a function in the usual way, that is, by looking to its history. I will call traits which meet these conditions members of the *historical trait type*. Let us now look more closely at these conditions.

Condition (a) picks out a class of intrinsic properties. What is required is a set which captures those properties which are possessed by all items to which a particular function would be ascribed, but are not possessed by items which would be ascribed a different function (namely, different traits), or no function at all (namely, severely malformed organic *stuff*). Here is a rough approximation of one way the etiolist might go in pinning down the nature of this set.³⁴ She could appeal to an appropriate range of instructions contained in the parts of the genetic code associated with recent ancestral items which performed F . So for any set of genetic instructions required to create a type of creature C , at a certain location in the genetic code will

33 My proposal here is not intended as conceptual analysis, I make no claim about the term 'function' as used by biologists, this in line with Millikan's programme (see for example Millikan 1984: 18, 1989b: 290, cf. Neander 1991a).

34 Thanks to Paul Noordhof for helping me work through the following approximation.

be instructions associated with recent ancestral items which performed *F*. To qualify as a member of a functional kind *K*, with the function of *F*-ing, a trait *t* must have been produced by instructions within a range considered a reasonable attempt to produce an item similar to ancestral items which performed *F*. This set of otherwise heterogeneous intrinsic structural properties cited in (a) then, can be picked out by an appeal to an appropriate range of allowed departures from genetic instructions associated with ancestral items which performed *F*. The idea here is that evolution chooses between intrinsic properties, function attribution is licensed when these intrinsic properties start to have relational functional properties, that is, when condition (b) is met.

For example, let us say that there is some set of instructions *i* present in the part of the genetic code associated with the production of ancestral human blood pumps. These instructions will produce an array of traits which will fall into one of three categories: *working*, *defective*, and what I will call *useless* (i.e., those which might be best described as ‘glob[s] of malformed organic matter’ (Millikan 1984: 25)). Our appeal to the set of intrinsic structural properties demarcates working and defective traits from useless traits, with those in the former two categories qualifying as members of the functional kind *heart*. Such categories will be roughly correlated with instructions which fall into the range considered a reasonable attempt to produce an item similar to ancestral items which pumped blood, and will group together everything we want to call a *heart*—working pumps and defective pumps.³⁵ So-called useless traits resulting from *i*—excluded from the functional kind—will correlate roughly with those instructions present in the part of the genetic code associated with the production of ancestral items which pumped blood, but which fall outside of the range of those instructions considered a reasonable attempt to produce such an item.

The idea then is that the set of intrinsic structural properties cited in (a) delineate a multiply realizable kind. The past instances of this kind are selected for in virtue of *F*-ing, which is adaptive. This is compatible with malfunction because not every way of realizing this kind needs to be in an *F*-ing contributing way.³⁶

Satisfaction of (a) is not sufficient for membership to a functional kind—this is in line with the historical theorist’s claim that an item’s history is necessary for function ascription. If possession of one of a set *S* of intrinsic structural properties in (a) was sufficient, depending on

35 This will of course admit of boundary cases (see fn. 40).

36 Thanks to Nick Jones for relaying my proposal back to me in these terms.

how we pinned down the nature of the set of intrinsic structural properties, we might get function ascription for first traits (those which are not descendents of *F*-ers with one of the properties causally responsible for the trait *F*-ing) and Swamp traits (those possessed by the spontaneously coalescing creatures of philosophical thought experiments). A token set of instructions being *associated* with the production of ancestral blood pumps requires that that token set of instructions *resembles* other token sets of instructions which have coded for the production of ancestral blood pumps. This means that condition (a) can be met by first traits and Swamp traits. When we step back and look at all the traits which were coded for by sets of instructions which produced blood pumps, first traits will be included in this set. As for Swamp traits, they too will meet condition (a) understood in this sense (due to Swampman being a ‘physical replica’ of a biological organism (Davidson 1987: 443)). So these traits may meet condition (a) (again depending on how we spelled out the nature of the set of intrinsic structural properties), but in their failing to meet condition (b), they do not qualify as members of a given functional kind. So they do not have functions, and thus they cannot malfunction.

A criticism which might be raised against my characterisation is that the requirement of a structural property is in tension with the etiological account, because proponents hold that *history is sufficient* for function ascription. It is true that proponents of the strong and weak etiological accounts have claimed that history is sufficient to determine purposiveness and norms of performance. However, I do not think that the appeal to a structural property would be rejected by etiological theorists.

In her definition of a higher-order reproductively established family, Millikan talks of ‘*similar* items’ and of traits being ‘*in some respects like*’ Normal members of a given reproductively established family (Millikan 1984: 24–5, my emphasis), I will outline what Millikan means by these terms in chapter five (§5.1), the point at the moment is to gesture at terms which indicate vagueness. She claims that the use of such terms reflects the vagueness of the question ‘whether a bit of matter should be called “a malformed eye” or merely “a glob of misplaced organic matter on the forehead”’ (Millikan 1984: 25). Also, in her definition of *heart*, Millikan speaks of hearts bearing ‘some resemblance to Normal hearts of that species’ (Millikan 1995d: 55, my emphasis). Similarly Price, a weak etiological theorist, speaks of two items belonging to the same functional type providing they were produced ‘*in more or less* the same way’ (Price 1995: 151, my emphasis). It looks then, like requiring a structural property for membership to a functional kind would be friendly to both strong and weak etiological accounts, and would help pin down more

precisely what Millikan³⁷ and Price³⁸ are getting at when they talk in terms of items being ‘similar’ or produced ‘in more or less the same way’.

Getting more specific about the kind of structural property required is tricky,³⁹ and though I have offered a way the strong etiolologist might go with regards to this, it is inevitable that vagueness will arise. Though I think vagueness of this kind is likely to infect any account of biological function owing to borderline cases, and so it does not speak against the conditions on function ascription I have offered.⁴⁰

Let us now look to (b). The three clauses in (b) are each doing important work, they make more specific the causal relationship held between a present day token and an ancestral token of a trait type, by requiring that the ancestral token satisfies clauses (i), (ii), and (iii). Let us first look to (b) [i]. Requiring the ancestral trait in the causal-historical relation to possess one of a set S of intrinsic structural properties, $\{p^1, p^2, p^3, \dots\}$ rules out the possibility of functional kind membership for traits which are descendents of something which performed F but which was a different kind of thing, for example, a blood-pumping pseudo-lung (I cannot call it a ‘lung’ because biological kinds are typed by their functions). The property in (b) [ii] is the success property, the property which allows the function to be performed, if a trait has this property it is a member of Davies’s *selected functional type*. This clause is important because there must be historical success for function attribution. Finally, clause (b) [iii] reflects the strong etiolologist’s commitment to the claim that variation is required for selection. That is, a trait has the function it does because the historical performance of that function accounts for the trait’s presence in

37 Millikan does not think that malformed devices are terribly important to the etiological theorist’s program (see fn. 30), though she thinks we could account for a malformed device using the conditions I introduced (Millikan: personal correspondence).

38 Price agrees that appealing only to history will not do, and that we need another (merely necessary, not sufficient) condition to meet Davies’s objection, though we should prioritize the historical condition (Price: personal correspondence). This is in line with what I have done. Condition (a) allows us to meet Davies’s objection, whilst condition (b) is still necessary (and allows us to specify the function of a given trait type).

39 So tricky we might be best leaving it to the anatomists. My aim here is not to give a foolproof pinning down of condition (a), but only to indicate that something like it is necessary (I am indebted to Paul Griffiths for discussion on this).

40 This is recognised—indeed embraced—by etiological theorists. Millikan claims that there ‘are lots of borderline cases of proper function, if not so many in nature, certainly in possible worlds’ (Millikan 2002: 115, fn. 2). When discussing the development of functional terms (‘Cummins biofunction’ and ‘proper biofunction’) she suggests that we should ‘not attempt to give these notions entirely clean boundaries. Nature has many important joints, *but these joints are seldom clean*’ (Millikan 2002: 122, my emphasis). Similarly Price claims that there ‘is nothing wrong with a certain amount of vagueness in our account of functions provided we are clear about what is going on’ (Price 2001: 29, fn. 19).

the current population, ‘as over against *historical* alternative traits no longer present’ (Millikan 1995c: 40). As I said earlier, a weak theorist could change this condition to ‘contributed—by *F*-ing—to some component of organismic fitness’ to reflect the fact that the weak account does not require *selection* over other traits.

Importantly, meeting the three clauses in (b) is not sufficient for membership to a functional kind. If it were this would be far too liberal—imagine the case in which something goes *so* wrong, that the trait produced is a so-called useless one. These items are not *malfunctioning* items, but rather not functional traits at all. Condition (a), the requirement of an intrinsic structural property, excludes such items from the functional kind. Also excluded from the functional kind are first traits and Swamp traits, though they may have the same intrinsic properties as other hearts, they do not have relational functional properties as they do not meet condition (b).

4.6 Responding to the Impossibility of Malfunction Objection

I am now in a position to respond to Davies’s objection. Before that, it is worth noting that the scope of Davies’s objection is not as wide as he takes it to be. The objection only goes through when the etiologist applies her account to congenitally damaged or diseased traits.⁴¹ That is, the etiological account is in trouble only with regard to its story about malfunctioning traits which are unable to perform their functions as a result of congenital factors. This is because conditions (i)–(v) refer to the very existence of a given token, and, as we have seen, for Davies, in order to be part of a given selected functional type, a trait has to possess the success property for being able to perform its proper function.

However, the set of congenitally damaged token traits does not capture all of the token traits which are considered malfunctioning. This is because there are cases where there is malfunction without congenital disease or defect.⁴² Davies claims, just before laying out his

41 There is of course a question about whether all of these kinds of cases will be due to abnormal conditions, rather than malfunction (see fn. 30). I suspect Neander would think that these cases are cases of malfunction, whereas Millikan would say that they are cases of mere failure to perform, due to conditions being abnormal for proper functional performance.

42 Or ‘dislocation’. In discussing possible malfunctions which may escape his objection, Davies considers malfunction due to dislocation and finds such cases to be unhelpful to the etiologist. A case of malfunction due to dislocation would be something like the following: suppose we remove a heart from a corpse and take it to be transplanted into another person. During this period, the heart fails to pump blood. It might be claimed that this is a malfunction of the heart, after all, it fails to do what it is supposed to do. Davies claims that this is the wrong way to think about the case, he says that the heart ‘is not malfunctioning’. The heart’s failure to pump blood, in this scenario, is not due to any change within the heart. It is due solely to dislocation from its normal environment’ (Davies 2000: 33). Davies is of course right, and he could have been stronger on this point. I think his use of ‘normal’ was in the statistical sense

objection, that ‘it is difficult to see what in the theory of selected functions or in the theory of weak etiological functions justifies the claim that functions persist even when the physical capacity is lost’ (Davies 2000: 20). Also, as we saw in his summary of the objection, tokens can ‘lack the defining property due to defect’ and they can ‘[lose] the defining property due to disease or damage’ (Davies 2000: 19). Here we see that Davies took his objection to apply to traits which malfunctioned, but did so not as a result of congenital disease or damage. Indeed, he claims that ‘an incapacitated token of T is *no longer* a member of the category of T’s’ (Davies 2001: 199, n. 6, my emphasis). Let us concede to Davies that congenitally damaged traits are not malfunctioning ones because they are excluded from the selected functional category. It does not follow that we have to accept that environmentally damaged traits are excluded from the selected functional category also, even though they were included before. The most we have to accept is the claim that it is a necessary condition on function possession that the trait in question *started* with the capacity to perform that function. To claim necessity for continued possession of that capacity is not something we can arrive at from Davies’s argument.

The etiological account does perfectly well then with respect to token traits which *had* the success property, and *were* able to perform their function, but now, due to damage or disease, are unable to do so (I will call these cases of malfunction *environmental* malfunction, as opposed to *congenital* malfunction). After all, these token traits *met* the condition put in place by the etiological account—they held the appropriate historical connection with previous reproductively successful token traits, and, also, these token traits *met* the condition put in place by Davies—they possessed the capacity to perform their function. They thus count as members of the selected functional type. If they *subsequently* fail to do what they are *supposed to do*, because they lose the property which granted them membership into the selected functional type, then they malfunction. This means then that traits which start as members of the selected functional type but are later unable to perform their function, can count as malfunctioning. These will be items such as hearts which *become* diseased and are unable to pump blood or pancreases which *become* unable to produce sufficient insulin.

(on pain of appealing to history to determine functional status), but the selected functions account gives another reason why conceiving of the heart as malfunctioning would be to make a mistake. That is by appeal to Normal conditions for the performance of the heart’s proper function. Being removed from the body and transplanted into another is not a Normal condition—that is, it is not a condition we need to appeal to in order to describe past functional performance (Millikan 1984: 33). And for this reason, it is a mistake to describe the heart in this case as malfunctioning. Concluding that it were malfunctioning would be like concluding that the human thumb malfunctions when it fails to grasp when the hand to which it belongs is tied. However, as I will now discuss, I think that there is another class of malfunctioning traits which escape Davies’s objection.

However, if the etiologist has only this to say, she can only account for normativity and function in the environmental case. Davies can grant that his objection does not work as applied to cases of environmental malfunction, but if this is all an account of function can accommodate, it leaves out an important class of malfunction, those which are congenital. Davies wants the etiological account to explain malfunction ‘due to defect, disease, or damage’, examples of which include ‘discoloration in wings, [and] hearts with damaged valves’ (Davies 2000: 33), and at least some of these kinds of malfunction will be congenital. Equally, etiologists in their discussion of malfunction cite examples which are plausibly—at least in some cases—congenital. Millikan talks of members of functional categories which are ‘diseased, malformed, injured, [and] broken’ (Millikan 1989c: 295), and Neander talks of items exhibiting ‘pathological deviations from the norm, due to disease, injury or deformity’ (Neander 1991b: 467). Millikan and Neander (among other etiologists) look to be⁴³ offering accounts which can accommodate environmental malfunction *and* congenital malfunction. Given all of this, Davies’s objection remains a serious one.

However, Davies’s objection can be answered if the etiologist accepts my characterisation, given in §4.5, of the causal-historical relation which holds between past and current tokens of a given trait type. I will first explain how the strong etiologist can respond to the objection, and then how the weak etiologist can too. Davies has it that the tokens of T cited in (v) have to refer to members of the selected functional type on pain of equivocation, as tokens of T cited in (i)–(iv) refer to members of the selected functional type (Davies 2000: 28). It is no good the strong etiologist meeting the equivocation charge by simply changing condition (v) so that it explicitly has ‘generic trait type T’ in ‘T’s place, because that would commit the account to something false. Let us take an example of Davies’s to bring out this point:

[c]onsider, for example, a distant ancestral population in which a precursor to the human heart first emerged. Suppose that, prior to the emergence of this proto-pump, nutrients were distributed and wastes collected by way of some sort of diffusion and secretion device. Relative to the generic trait type ‘circulatory mechanism’ the population contained two morphologically distinct variants, proto-pumps and diffusion/secretion devices, and the difference between these was selectively significant. (Davies 2000: 26)

43 In Millikan’s case, this is no longer what she is up to (see fns. 30 and 37).

If conditions (i)–(v) were applied within this story, and the strong etiologist took tokens of T in condition (v) to apply to the generic trait type, she would get a very unhappy result indeed. This is because a generic trait type includes all items that were apt for selection, whilst a selected functional type includes only members with the property selected for. So, with this example, the result of the selective success of the proto-pump, would be the persistence or proliferation of members of the generic trait type *circulatory mechanism*. This would include tokens of proto-pumps *and* diffusion/secretion devices which performed less well against the proto-pump and so were not selected for. The strong etiologist does not want to commit herself to the selective success of proto-pumps being responsible for the proliferation of diffusion/secretion devices, quite the opposite in fact. So, Davies claimed that the strong etiologist has to say tokens of T in (v) refer to members of the selected functional type, but by doing so, she loses her claim to an account of malfunction.

If the strong etiologist adopts my proposal, she can now meet Davies’s objection by reading ‘tokens of T’ in conditions (i)–(v) as follows:

- (i) Past instances of tokens of the selected functional type in O performed F in E,
- (ii) The selected functional type was heritable,
- (iii) Past performances of F caused an increase in O’s relative ability to satisfy demands of E (relative to other organisms in the population not members of the selected functional type),
- (iv) This increase in O’s ability to satisfy selective demands of E resulted in an increase in O’s long-term relative rate of reproduction,
- (v) This increase in relative reproduction resulted in the persistence or proliferation of O and hence tokens of the *historical trait type*.

The strong etiological account now has it that the success of members of the selected functional type resulted in tokens of the historical trait type. And tokens of the historical trait type can malfunction, because (a) they possess one of a set of intrinsic structural properties, and (b) they are governed by the norms of performance bestowed on them by their historical connection to prior successful tokens of the generic trait type to which they belong.

There is a worry one might have at this point. The proposal is to extend attribution of function to a class of items which includes members of the selected functional type, but is not as extensive as to include all members of the generic trait type. Why is this legitimate, but attributing functions to members of the generic trait type which meet (b) is not? That is, why

does the meeting of condition (a), *in addition to* meeting condition (b), legitimize function ascription? The way I explicated the property which is one of a set of intrinsic structural properties, cited in (a) can answer this worry: the genetic instructions which, for example, code for a *blood pump* are not coding for the production of an item which falls under the generic trait type (they are not, say, coding for the production of a blood pump or a diffusion/secretion device). Rather, it is their coding for the more specific thing—a blood pump—which is relevant. They are coding for a blood pump, but, due to variation, can producing both working and defective blood pumps. It is this more precise characterisation that makes legitimate the attribution of function to traits which meet conditions (a) and (b). As long as a trait is part of the historical trait type, then the norms of performance apply to it, and it can malfunction. The strong etiological theorist's account can then, contra Davies, accommodate malfunctions.

The weak etiological theorist can give the same response, but membership to their *historical trait type* will be easier because of the more liberal condition (b) [iii]. The weak etiological account now has it that the success of those tokens which contributed to organismic fitness resulted in tokens which are members of the historical trait type. And, once again, tokens of the historical trait type can malfunction, because (a) they possess one of a set of intrinsic structural properties, and (b) they are governed by norms of performance bestowed on them by their historical connection to prior tokens of the generic trait type to which they belong which contributed to organismic fitness. If a trait token is a member of the historical trait type, then it can malfunction in virtue of there being norms of performance which apply to it. The weak etiological account can also accommodate malfunctions.

Davies's objection from the impossibility of selected malfunction teaches us something important. Etiologists need to pin down the causal-historical relationship held between present-day tokens and ancestral tokens of a given trait type in order to accommodate malfunctioning tokens. I proposed a way in which this might be done in §4.5, by suggesting two conditions on membership to a functional kind. If the etiologist adopts these conditions, she does not have to accept that functional kinds are defined by reference to successful ancestral tokens, it was this which made her account vulnerable to Davies's objection. Rather she can hold on to the historical component of her account—it is after all by reference to a token trait's history that a specific function can be specified—and, by adding the requirement of a structural property to the conditions on membership to a functional kind, she can distinguish malfunctioning tokens of a given type from so-called useless ones. I conclude then that Davies' challenge can be met. The etiological account can ground a notion of biological

normativity and etiological malfunctions are possible.

4.7 The Circularity Objection

The last thing I will do in this chapter is consider another objection which can be responded to given my proposal. Nanay claims that ‘there is no coherent, noncircular way of individuating trait types that is available to the etiological theory of function’ (Nanay 2010: 415). Nanay’s objection here, if it works, will work equally against both strong and weak etiological accounts. Nanay offers three ways of individuating trait types, and claims that the etiolologist cannot help herself to any of them. His wider goal is to propose a modal theory of function which does away with trait type individuation and attributes function to token traits by appeal to possible worlds. I will not discuss his alternative theory here because my proposal can answer his objection to etiological accounts.

The first way the etiolologist might go in individuating trait types is by what Nanay calls the ‘Functional Criteria’, which has it that ‘[a] token object belongs to a trait type *T* if and only if it has certain functional properties: if it has the function to do *F*’ (Nanay 2010: 415). However, this strategy will lead the etiolologist into circularity, for if she adopts it, she thereby uses the notion of function to individuate trait types, when it is the explanation of function which is at stake. As Nanay puts it:

the claim that *X** (the trait whose function we are explaining) is a token of type *X* (the traits that have been selected in the past) is part of the *explanans*. Hence, we cannot use the *explanandum* (function) to explain part of the *explanans* (why *X** is a token of type *X*). (Nanay 2010: 416)

The next option is to individuate trait types by appeal to morphological criteria, so that ‘a token object belongs to a trait type *T* if and only if it has certain morphological properties’ (Nanay 2010: 416). This strategy is obviously hopeless and would be immediately eschewed by any historical theorist, we saw earlier that membership to a biological kind is not determined by properties of a given item, but by its history. If we adopt the morphological criteria, we lose any hope of giving an account of malfunction, because a heart that is malformed may not have the right morphological properties, but it is nevertheless a heart (Nanay 2010: 416).

The final option is to individuate trait types by appeal to homological criteria, so that ‘what guarantees that two traits are tokens of the same type is that they are homologues: they have common descent; they are members of the same “reproductively established family”’ (Nanay 2010: 417). However, this strategy is also no good for the etiolologist, Nanay gives the following example to illustrate why:

[t]he forelimbs of vertebrates, such as the wings of birds, and the forelegs of ancient amphibians are homologous: the wings of eagles belong to the same reproductively established family as the forelegs of some ancient amphibians. According to the suggestion for individuating trait types we are considering here, they must be of the same trait type. (Nanay 2010: 417)

But of course these two traits are not tokens of the same type, and so this strategy will give the etiologist the wrong result. Nanay concludes that the etiological account cannot adopt any of the three possible ways of individuating trait types, and so it ought to be rejected (Nanay 2010: 419).

The etiologist can respond to Nanay's objection by adopting my proposal.⁴⁴ Here, as with Neander and Alex Rosenberg's response (see fn. 44), the claim is that the etiologist does not need to type traits before she can attribute functions, rather, 'trait individuation and function attributions supervene on the same underlying facts' (Neander and Rosenberg 2012: 617). Those facts being the ones in conditions (a) and (b). There is no circularity for my proposal, because both trait type individuation and function ascription are grounded on history and possession of one of a set of intrinsic structural properties.

Perhaps my condition (a) is vulnerable to Nanay's objection to defining trait types by appeal to morphological criteria. The problem here was that an appeal to morphological criteria ruled out the possibility of malfunction. However, as we saw in my response to Davies, my proposal is not guilty of this. Rather, condition (a) captures those properties which are possessed by all items to which the etiologist would ascribe a particular function (and have fall into a particular functional kind), but are not possessed by items which she would ascribe a different function (or have fall into a different functional kind), or no function at all (and thus not qualify as a token of any trait type). So Nanay's claim that 'a malformed heart [...] is a heart all the

44 Neander and Rosenberg have responded to Nanay in defence of the etiological account. They argue that the etiological account can appeal to lineages: 'a lineage of traits is segmented (or as we say, parsed) both for the purpose of classifying traits and for the purpose of ascribing functions by changes in the selection pressures that have operated on it. No circularity is involved' (Neander and Rosenberg 2012: 618). The claim here is that the etiologist does not first need to type traits before she identifies the function of a given item. Rather, both trait individuation and function attribution co-supervene 'on the same underlying facts about trait lineages and the selection pressures operating on them' (Neander and Rosenberg 2012: 417). However, Nanay claims that this line will not work because we cannot identify lineages without referring to trait types, and so circularity sneaks back in (Nanay 2012: 624). I will not discuss Neander and Rosenberg's response and whether it works any further, as I think my proposal can meet the objection.

same' is accommodated by my proposal.

4.8 Conclusion

In this chapter I developed and defended the etiological account of biological function. I proposed a version of the account which made clearer the causal-historical relation held between ancestral and present day tokens of a trait type. This proposal allows the etiological account to ground a notion of biological normativity, and respond to Davies's Impossibility of Malfunction Objection. It also enables the etiological account to respond to the Circularity Objection raised by Nanay. I will now adopt the etiological account in the following chapters in which I formulate my account of belief.

A BIOLOGICAL ACCOUNT OF BELIEF

In this chapter I will outline the first part of my account of belief and discuss the explanatory work it can do. First I will define some functional terms before ascribing the function of producing true beliefs to our belief-producing mechanisms. Note that I have moved here from talking about *beliefs* to the mechanisms which produce them. This is because my claim will be that the features which need explaining simply result from the ways in which beliefs are produced. I will then show how our mechanisms for belief-production having this function, given the essential nature of the states they produce, can give an account of the link between belief and truth. On my account, the features of belief I seek to explain are *contingent* features, grounded in the biological histories of our belief-producing mechanisms.

5.1 Functional Terms

In chapter one I defended the motivational account of belief. However, I also argued that the motivational account of belief cannot explain the existence of the three features of belief which indicate a link between belief and truth. I suggest that a biological account of belief can be the add-on required to provide these explanations. Claims about the biological functions of our mechanisms for belief-production and the products thereof can fill in these details and meet the belief theorist's explanatory burden. On my account the motivational role of belief defended in chapter one is taken to provide a necessary and sufficient condition for a state's being a belief, whilst the features of belief which indicate a link between belief and truth are contingent features of the beliefs of *some believers*.

We need to distinguish between *Weak* and *Strong* Epistemic Normativity. The first can be explained by appeal to the function of producing true beliefs. However, if we want an explanation of the second, we need to appeal to another function. The stronger version of Epistemic Normativity is *not* a feature which links belief to truth. For this reason, I save discussion of Epistemic Normativity for the next chapter. In this chapter I will discuss my account with respect to its explanations of Transparency and the Uncontrollability Thesis.

In order to describe my account, I need to lay some definitional ground. Firstly, as discussed in chapter four, I adopt the etiological account of biological function. I have developed and defended this approach to biological function arguing that the conditions on function ascription had better include possession of one of a set of intrinsic structural properties, in addition to the appropriate history. I claimed that requiring possession of a

structural property would help pin down what etilogists are getting at when they talk in terms of items being ‘similar’ (Millikan 1984: 24) or produced ‘in more or less the same way’ (Price 1995: 151), as well as excluding what I called ‘useless’ traits from a given functional kind. I think this development of the etiological account complements, rather than opposes, the account of function Millikan develops. Given this, I will help myself to some of her terms, but such terms are to be understood both as Millikan defines them, and as additionally fitting into my developed account (i.e., add to the conditions required to have a particular function my conditions (a) and (b) (§4.5)). Millikan, as a strong etiological theorist, thinks that selection over alternative traits is required for function ascription. However, the terms I will take from her could equally be given a weak reading (for example, change talk of effects which were *selected for* to talk of effects which *contributed to some component of organismic fitness*).

I will first outline several definitions of terms introduced by Millikan, which will be used in my account of belief. Firstly, for a trait to have a *proper function* it needs to have

originated as a “reproduction”⁴⁵ [...] of some prior item or items that, *due* in part to possession of the properties reproduced, have actually performed [that function] in the past, and [the trait] exists because (causally historically because) of this or these performances. (Millikan 1989c: 288)

When we ascribe to a trait a proper function, we ascribe to it a function that was performed by⁴⁶ its ancestors, which ‘helped account for the proliferation of the genes responsible for it’ and thus also ‘helped account for its own existence’ (Millikan 1989c: 289). The heart has the proper function of pumping blood, because it originated as a reproduction of an ancestral heart, which performed the function of pumping blood in the past, and it is this fact which accounts for the proliferation of the genes responsible for the heart, and therefore the existence of it. Millikan’s definition of *proper function* here meets clauses [ii] and [iii] in (b). For a trait to have a *proper function* in my terms then, we need only add to Millikan’s definition, that it possesses one of a set of intrinsic structural properties (condition (a)), and that the trait to which it stands in a causal-historical relation, also possesses one of a set of intrinsic structural properties ((b) [i]).

45 Millikan sets three conditions on what it takes for there to be a case of reproduction. In summary—and very roughly—“for B to be a reproduction of A it is necessary only that there be some way of describing B’s causal history, holding certain conditions constant [...] such that it is explained why B had to be like A, whatever the character of A, within a certain range of character variation” (Millikan 1984: 20).

46 Though not in all cases possessed by these ancestors. First items which performed a given function, did not possess that function, as they did not meet condition (b).

This notion of proper function can then be broken down into several sub-types, some of which I define here. The first sub-type of a proper function is a *relational proper function*. To say of a device that it has a relational proper function is to say of it that ‘its function [is to] produce something that bears a specific relation to something else’ (Millikan 1984: 39). Relational proper functions are the effects which have helped account for the selection of the producing mechanism (Millikan 1984: 26). Consider the chameleon’s mechanisms for changing the skin pigmentation. These mechanisms have the relational proper function of producing a skin pattern which matches the chameleon’s environment⁴⁷ (Millikan 1984: 39).

Related to this, devices which have a relational proper function also have—at any time of functional performance—*adapted proper functions*. These are functions which the device has ‘given some specific thing that [it] is now supposed to produce in relation to’ (Millikan 1984: 40). In the case of a chameleon, when placed on a green leaf, the mechanisms for skin pigmentation acquire the adapted proper function of changing the skin green. The devices produced by such mechanisms are *adapted* devices, which bear a relation to something else. That which the adapted device is adapted *to* is the *adaptor* (Millikan 1984: 40). In the case of the chameleon’s mechanisms for skin pigmentation, the produced pattern in this instance is the adapted device, with the adaptor being the leaf which the device is produced in relation to.

Adapted devices do not have proper functions, as they do not meet the conditions thereon. They have not ‘originated as a “reproduction”’ (Millikan 1989c: 288) of prior adapted devices, and their existence is not due to the functional performances of prior adapted devices. Rather, adapted devices have *derived proper functions*. For a trait to have a *derived proper function* it needs to have ‘originated as the product of some prior device that, given its circumstances, had performance of [that function] as a proper function and that, under those circumstances, normally causes [that function] to be performed by *means* of producing an item like [that trait]’ (Millikan 1989c: 288). An example of an item with an invariant⁴⁸ derived proper function is the pattern of a chameleon at any given time. The pattern ‘has an invariant derived proper function

47 Recent work on African dwarf chameleons has suggested that the capacity for colour change is actually associated with social signals which are striking to chameleons’ visual systems. Also, the idea that the capacity for colour change is linked to camouflage has been undermined by research which suggests that ‘capacity for colour change is unrelated to the variation in the environmental backgrounds that chameleons must match in order to be camouflaged’ (Stuart-Fox and Moussalli 2008: 23). I use this example only illustratively, that is, I use it to demonstrate the kind of mechanism which has Millikanian relational proper functions. Nothing in my account hangs on the biological validity of this example.

48 We can also talk about derived proper functions which are variant. In the case of the chameleon on a green leaf, its skin pattern at that time has the variant derived proper function of being green—variant functions change depending on the adaptor in each case.

to make the chameleon invisible to predators, hence to prevent it from being eaten' (Millikan 1984: 42).

The final term which I will put to use is 'Normal' (note the capitalisation). I will follow Millikan is using 'Normal' in a normative, historical—as opposed to statistical—sense. A *Normal explanation*, for example, is one which 'explains the performance of a particular function, telling how it was (typically) historically performed on those (perhaps rare) occasions when it was properly performed' (Millikan 1989b: 284). A *Normal condition* for the performance of a given function is one 'the presence of which must be mentioned in giving a full Normal explanation for performance of that function' (Millikan 1989b: 285).

Proper functions can only be had by items which are members of what Millikan calls *Reproductively Established Families*, of which there are two types, *First order* and *Higher order*. First order reproductively established families are sets of entities which have 'the same or similar reproductively established characters derived by repetitive reproductions from the same character of the same model or models' (Millikan 1984: 23). If an item has in common certain properties or a certain character, with other members of a reproductively established family, as a result of reproduction, that item is a member of a first order reproductively established family (Davies 1994: 366). Higher order reproductively established families are constituted by sets of 'similar items produced by members of the same reproductively established family, when it is a direct proper function of the family to produce such items and these are all produced in accordance with Normal explanations' (Millikan 1984: 24). An item is a member of a higher order reproductively established family then only if it 'shares certain properties in common with other members as a consequence of being produced (not reproduced) by members of some specific first-order reproductively established family, where members of this first-order family have the direct proper function of producing members in the higher-order family' (Davies 1994: 366). The vagueness which appears in Millikan's definitions here ('similar items', 'shares certain properties') can be resolved by appeal to possession of one of a set of intrinsic structural properties in my condition (a) (§4.5).

With the relevant terms defined, I move on to offer my functional claim about belief.⁴⁹ It is

49 Millikan has made some of the claims about belief which I make, she has not, however, put them to work in the same way I do. She claims, for example, that it is a 'function of our belief-fixing systems to fix true beliefs' (Millikan 1995a: 177, see also 1995b: 243–4). She ascribes to belief several functions, including '[participating] in inferences in such a manner as to help produce fulfilment of desires' and '[participating] in inferences to yield other [true] beliefs' (Millikan 1995d: 71). As we shall see, I ascribe only two functions to belief. Not because I think that Millikan is wrong in her multiple ascriptions, but because I am interested in how much work an appeal to biological

important to bear in mind that I am building a biological account of belief onto a motivational approach which was defended in chapter one. What I think belief *essentially* is, is a cognitive state which ‘*by itself*, and relative to a fixed background of desires, disposes the subject to behave in ways that would promote the satisfaction of his desires if its content were true’ (O’Brien 2005: 56), *across contexts*. A biological account of belief, considered as this sort of state, is going to be able to do the explanatory work called for by the features of belief which indicate a link between belief and truth. Such an account will ascribe a function to our mechanisms for belief-production, which will allow for an explanation of why they behave as they do, that is, the constraints under which they operate.

I claim that the mechanisms which produce beliefs in us have the relational proper function of producing *true beliefs*. I adopt a kind of Evolutionary Reliabilism which is the claim that ‘natural selection tends to favour *reliable*⁵⁰ belief-producing mechanisms’ (Ramsey 2002: 16, my emphasis).

In this chapter I will claim that our mechanisms for belief-production having this function explains the presence of Transparency and the Uncontrollability Thesis. In the next chapter I will argue that we must recognise a second function proper to our belief forming mechanisms, and it is both this function ascription and the function of producing true beliefs which explains Epistemic Normativity.

5.2 Function: Production of True Beliefs

I claim that the mechanisms which produce beliefs in us have the production of *true beliefs* as one of their proper functions, which is a *relational* proper function. Such mechanisms are members of a higher-order reproductively established family in the same way that hearts, livers, lungs, and so on, belong to higher-order reproductively established families. With respect to *her* heart for example, Millikan claims that it was ‘produced under Normal conditions in accordance with the proper functions of certain of [her] genes which were directly copied from [her] parents’ genes’ (Millikan 1984: 25, emphasis removed). We can give a similar account of our belief-producing

function can do in explaining the relationship between belief and truth. My relatively limited function ascriptions should be seen to reflect my relatively limited project.

⁵⁰ Ramsey understands reliability in terms of truth. He claims that on an evolutionary reliabilism view, ‘the truthfulness of our beliefs actually helps explain their adaptiveness [...] the accuracy of [cognitive states], in turn, makes them more inclined to generate fitness-enhancing behavior’ (Ramsey 2002: 19). However, I want to understand reliability not in terms of truth, but in terms of the generation of fitness-enhancing action.

mechanisms; that is, it is a direct proper function of certain genes (themselves members of a first-order reproductively established family) to produce our belief-producing mechanisms—which are similar to each other—and are produced in accordance with Normal explanations (Millikan 1984: 24).

The claim that the mechanisms which produce beliefs in us have the production of true beliefs as one of their proper functions relies on two claims:

(F1) Our mechanisms for belief-production were selected for.

(F2) Our mechanisms for belief-production were selected for producing true beliefs.

I will argue for each of these claims in turn.

5.2.1 (F1) Our Mechanisms for Belief-Production were Selected For

Claim one could instead be the weaker claim that our mechanisms for belief-production contributed to some component of organismic fitness, if one were a weak etiological theorist. As I said in chapter four (§4.3), I think that the strong account is correct, but note that the weak account can do the work required of the belief theorist. I will continue to speak in terms of the strong account, but everything I say can be translated into the terms of the weak account, for example, instead of saying that something was selected for (strong), read this instead as something's contributing to some component of organismic fitness (weak). To argue for claim (F1) is to argue for the appropriateness of a biological account of the nature of our mechanisms for belief-production. Given that I adopt the etiological account of biological function, if a device was not selected for that device does not have a function. So our mechanisms for belief-production need to have been selected for in order for them to be functional devices. I will give three reasons for thinking that this is the case. The first is stated by Millikan when she claims that

[b]eing “built” by natural selection is sufficient for proper function, being maintained by natural selection is independently sufficient for proper function. It is hard to see how modern human cognition could fail to be caught in one or more of these nets. (Millikan 1995c: 49)

The point here is that modern human cognition cannot fail to meet at least one of the sufficient conditions for biological function possession on the etiological account. An extension of this would be the claim that our mechanisms for belief-production, *as* a part of human cognition,

can equally not fail to meet one of these conditions, and so they too are functional items. This in turn supports Millikan's claim that 'psychological classification is biological classification, [...] categories such as belief, desire, memory, percept, and purposive behavior are biological function categories' (Millikan 1995a: 172).

The second reason for thinking that our belief-producing mechanisms were selected for relates to the huge biological costs of producing and maintaining them. Such costs may be related to the long period of infant dependency required for the development of the human brain, as well as the large energy costs of running it. Evan Fales takes this line claiming that

beliefs must be generated – in our case, by distinct mechanisms linked to the dozen or more sensory, proprioceptive, and introspective modalities with which we are equipped. Next, to be of use, they must be catalogued and stored for efficient retrieval, they must be constantly squared with one another to insure as much consistency and inductive coherence as possible, and there must be inferential mechanisms devoted to the production of further beliefs. (Fales 2002: 48–9)

Fales claims that the existence of such biologically expensive capacities provides a *prima facie* reason for 'assigning a low probability to the development of such mechanisms unless they confer a decided selective advantage' (Fales 2002: 49). This is not conclusive of course, it certainly does not *prove* that our mechanisms were selected for (as opposed to their being the result of selection *of* or their being a spandrel⁵¹), but it certainly lends support to that claim.

A third reason to think that our belief-producing mechanisms have been selected for is that the formation of beliefs produces changes to a creature's dispositions which alter its responses to the environment. Given that beliefs are essentially motivational devices (as argued in chapter one), adaptive pressures will be in play. The formation of beliefs produces changes in the way the subject navigates within and responds to its environment, and so the mechanisms which produce them will be sensitive to adaptive pressures. If this is right, the biological costliness of our mechanisms for belief-production as mentioned above is less important. Even if beliefs were really biologically cheap to produce, it would still be to an evolutionary disadvantage for

51 I leave aside the possibility that our mechanisms for belief-production are a spandrel, a by-product of other devices which were selected. Millikan hints at this kind of approach, and finds it wanting: '[t]here must [...] be a finite number of general principles that govern the activities of our various cognitive-state-making and cognitive-state-using mechanisms and there must be explanations of why these principles have historically worked to aid our survival. To suppose otherwise is to suppose that our cognitive life is an accidental epiphenomenal cloud hovering over mechanisms that evolution devised with other things in mind' (Millikan 1995d: 59).

the mechanisms to produce false ones (see §5.2.2) as they affect how a creature navigates within, and responds to, its environment.

I have outlined three reasons for thinking that it is appropriate to give a biological account of our mechanisms for belief-production in line with the etiological account of biological function. Firstly, human cognition, of which these mechanisms are a part, looks like the kind of thing which cannot fail to meet at least one sufficient condition on biological function possession. Second, given the huge complexity and associated biological costs of producing and maintaining these mechanisms, we have a *prima facie* reason for thinking that they were selected for. Third, beliefs are essentially action guiding and so vulnerable to selective pressures. So let us say that our mechanisms for belief-production were selected for, the task now is to show that they were selected for the production of *true* beliefs.

5.2.2 (F2) Our Mechanisms for Belief-Production were Selected for Producing True Beliefs: Pre-Emptying Worries

Claim (F2) is that our mechanisms for belief-production were selected for producing true beliefs. The first thing to do in arguing for this claim is to show that true beliefs are adaptive, since something can not be selected *for*⁵² unless it is adaptive, unless it confers a selective advantage. Given the motivational profile of belief I argued for in chapter one (§1.2), beliefs being true will have obvious biological advantages. Beliefs which are true are more likely to dispose a creature to act in ways which will satisfy its desires, be they for food, warmth, sex, and so on. The claim that true beliefs are adaptive has been taken to be obvious by many philosophers, W.V.O Quine for example claims that ‘creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind’ (Quine 1985: 39). Similarly, Evan Fales claims that ‘[*true* premises guarantee true conclusions: so a system that relies consistently upon true inputs to guide inference and action can employ general rules and hope to get things (i.e., action) right’ (Fales 2002: 51), and Dennett claims that ‘[n]atural selection guarantees that *most* of an organism’s beliefs will be true’ (Dennett 1987: 75). Finally, from Michael Lynch:

imagine what would happen if I were to *constantly* form false beliefs in the real world. Walking out of my door, I would fall down [...] I might put my car in drive when I meant for it to reverse [...] or

52 Not to be confused with selection *of* (see Sober 1984: 99–100).

think safe food was poisonous and poisonous food safe. (Lynch 2004: 48, cited in Hazlett 2013: 57)

However, other philosophers have argued that the assumption that true beliefs are adaptive is unwarranted. Christopher Stephens claims that ‘there is no monolithic answer to our question about the evolutionary value of believing the true’ (Stephens 2001: 162). He demonstrates the problem with the following thought:

suppose there is only a small chance that a tiger is nearby. It is better to be safe and believe that *a tiger is nearby* rather than believe that *no tiger is nearby* and be sorry if it turns out you’re wrong. (Stephens 2001: 162)

This way of setting up the problem does not constitute an objection to the claim that our mechanisms for belief-production were selected for producing true beliefs. It certainly does not go very far in demonstrating that true beliefs—or the mechanisms which produce true beliefs—are not adaptive (and if this were right then the mechanisms could not have been selected *for*). To see this suppose the following scenario holds: the mechanisms in us which produce beliefs have the function to produce true beliefs, and they are *completely infallible in their doing so*. They never fail to produce a true belief in the appropriate circumstances, and they never get things wrong such that a false belief is produced. Now if this were the case, the way Stephens has set up the better-safe-than-sorry objection is such that it cannot get any grip. If my mechanisms for belief-production really were infallible in this way, they would ensure my safety in cases like Stephens’s, without my being sorry. It is not safer for me to believe that a tiger is nearby when there is not one—even though there is a high cost of my being wrong, because, having infallible belief-producing mechanisms, I would never get it wrong. Having true beliefs in cases like these is never maladaptive, only false beliefs are. It is not relevant to this point that *our* belief-producing mechanisms are not infallible in this way. Given that Stephens wants to argue for the claim that true beliefs are not always adaptive, his appealing to the possibility of taking a false belief to be true is not helpful to this pursuit. He needs the extra premise that we have fallible mechanisms for belief-production. This is a premise I can of course grant, but it needs to be explicit in the argument. So, given our fallibility, sometimes it is better to have a false belief and be safe.

A better way to set up the problem comes from Stephen Stich—he offers the example of Harry who has a true belief that his flight leaves at 7:45. Having this belief contributes to Harry’s being on time to make the flight. Unfortunately for Harry, the aeroplane crashes and he dies. Regarding this, Stich draws the following moral:

in the case of Harry it is clear that if he had had one more false belief and one fewer true one, and if everything else in his cognitive life had remained as much the same as possible, his life would have been longer. (Stich 1990: 123)

Clearly in this case Harry's having a true belief was not good for him. His having a false belief—that his flight left an hour later, would have been more adaptive. Stich dramatises this thought by considering the adaptiveness of having TRUE**** beliefs, rather than true ones. We are to consider a case of non-standard mapping of mental states to propositions. This case is precisely like the standard mapping except for one thing: '[t]he belief that Harry would express by saying '[m]y flight leaves at 7:45' is mapped to the proposition that his flight leaves at 8:45, and the belief Harry would express by saying '[m]y flight leaves at 8:45' is mapped to the proposition that his flight leaves at 7:45 (Stich 1990: 123). Here mental states are mapped to their TRUTH**** conditions. We learn from the case of Harry that sometimes 'having TRUE**** beliefs is more conducive to survival than having true ones' (Stich 1990: 123).

On the wider question of the adaptiveness of true belief in general, Richard Feldman claims that the adaptiveness of true beliefs is far from being established. He thinks that 'the adaptiveness of believing the truth is not something that can be assessed independently of other factors' (Feldman 1988b: 225). This is because beliefs—whether true or false—only impact on the survival of a creature insofar as that creature uses its beliefs as a basis for action. Or, if the desires of the creature were destructive ones, true beliefs may well be maladaptive. Feldman claims that the question of whether true beliefs are adaptive 'depends on what one will do on the basis of those beliefs' (Feldman 1988b: 225).

Here is what we have learnt from the above considerations:

- (C1) True beliefs are not always adaptive.
- (C2) Selection of mechanisms which produce true beliefs is not the *optimal* adaptive strategy.

I will consider these claims in more detail now, arguing that their truth is not problematic for my account.

5.2.2.1 (C1) True Beliefs are Not Always Adaptive

The claim that true beliefs are not always adaptive is supported by cases like that of Harry, in which having a true belief has disastrous consequences for the believer. A belief that gets one killed is no less maladaptive just because it is true. Stich asks: 'Is true belief *always* more

conducive to survival than false belief? Clearly the answer is no' (Stich 1990: 122–3). Evidently, this is right.

Also, consider beliefs which are not obviously related to survival—those beliefs which look to have no adaptive value. This appeal to sophisticated or useless true beliefs to support the claim that true beliefs are not always adaptive has its analogue in a similar move made against teleosemantic theories of content. The teleosemantics program provides a naturalistic account of mental representation by appeal to biological function; truth conditions of intentional states are given in terms of what those states are *supposed* to do (Macdonald and Papineau 2006: 1). An objector to this account might claim that the teleosemantic approach is plausible with respect to determining the content of states which have an impact on a creature's fitness—so states concerned with food, warmth, sex, and so on. However, there is no reason to think the teleosemantic account can give content ascriptions to states which do not or cannot impact on fitness (Neander 2012). If the content of beliefs, say, can be determined by what those beliefs are *supposed* to do, how is a teleosemanticist going to ascribe content to my beliefs about quarks and wave functions? Equally, one might think that my account of belief is going to do some good work when we are considering cases of beliefs which are relevant to a creature's survival—beliefs about the location of food, shelter, and so on. Clearly, *those* beliefs being true will be adaptive. However, my account is perhaps less convincing when we consider non-survival directed beliefs; there is no reason to think that *those* beliefs being true would be adaptive. There is no reason to think that *those* beliefs have the derived proper function of having true contents, that they are *supposed to be* true. As Feldman puts the point:

[w]hile it is probably true that accurate beliefs about one's food and one's predators are beneficial, it is also true that there are many cases in which the truth or falsity of our beliefs is completely irrelevant to the survival of our species. Does it really matter if we have true beliefs about the stars of the situation comedies we watched on TV in the 1950s? (Feldman 1988b: 226)

Clearly, if my account denied the claim that true beliefs are not always adaptive, it would be in serious trouble. Fortunately, it does not. I do not need to deny that true beliefs are not always adaptive in order to claim that beliefs have the derived proper function of having true contents. Sometimes adaptive traits, in performing their proper function, are not adaptive, or even *maladaptive*. The blinking mechanism of the eye will be maladaptive if one's victory over a conspecific is decided by a staring competition. Or if one blinks so much to remove debris that

one misses a predator.⁵³ But that is not to say that there is anything wrong with ascribing to this mechanism the function of keeping the eye moist. Pointing to an abnormal case in which the mechanism's performance of its proper function is maladaptive is not to provide reason for thinking that it does not have that function. Though these examples are contrived, they are no more so than Stich's case of Harry.

As for cases in which it does us no good to have a true belief, once again, this does not speak against the functional claim. We can think of traits of biological organisms which, sometimes, their performing their function is not particularly useful. Function ascriptions are general claims, they do not apply to individual instances of functional performance.⁵⁴ Consider a chameleon in captivity. Its mechanisms for changing skin pigmentation perform their relational proper function by changing the chameleon's skin so as to match its immediate environment. Is its doing so conferring a benefit on the chameleon? Probably not, given that there are no predators. However, it remains the case that these mechanisms have the relational proper function of producing skin patterns which match the chameleon's environment. And it is equally the case that the produced devices have the derived proper function of matching the environment. In this case, the chameleon finds itself in abnormal circumstances, but that does not speak against its mechanisms for skin pigmentation having a relational proper function.

Also, there is some room in my account to accommodate Feldman's worry. Though our deliberative belief formation is always transparent to truth considerations, we require various amounts of evidence:

if I am deliberating about whether to believe that *The Simpsons* will be on TV tonight, I might be entirely satisfied by relatively minimal evidence, and thus form a belief. (Steglich-Petersen 2006: 503)

So though our mechanisms for belief-production produce beliefs which may not be very useful in terms of biological adaptiveness, it does seem to be the case that when they do so, there is some sensitivity to this. Indeed,

it would seem to be no accident but a wise choice of evolutionary design that our dispositional beliefs come in varying strengths. To waste time checking over and over p when it doesn't really

53 Thanks to Jane Tomlinson for this example.

54 The exception to this is variant derived proper functions. These functions are ascribed to adapted devices—devices which have relational proper functions—and take into account the particular performance of the relational proper function (see fn. 48).

matter that much whether p would be an inefficient design. But to fail to check whether p when it really matters a lot whether p would be disastrous design. (Millikan 1995b: 251)

Recall my discussion of confidence thresholds in chapter two (§2.3.1.2), non-epistemic considerations can change what a subject requires as sufficient evidence to form a belief. When deliberating on whether *The Simpsons* will be on, as in the case above, I require much less evidence to form the belief than I would if I were deliberating on whether or not the bank is open on Saturday when I have an important cash to cheque by Monday. So though true beliefs are not always adaptive, our mechanisms for belief-production are sensitive to this insofar as when we are not in a situation in which we really need to get things right, we are more easily satisfied that we meet the requirements on believing that p .

The overarching point of this section is the following: of course true beliefs are not always adaptive, but nor do they need to be for it to be the case that our mechanisms for belief-production were selected for producing true beliefs. It only needs to be the case that at key moments in the evolutionary past, mechanisms for belief-production conferred a selective advantage when they produced true beliefs. The fact that true beliefs are not always adaptive does not speak against this claim.

5.2.2.2 (C2) Selection of Mechanisms which Produce True Beliefs is Not the Optimal Adaptive Strategy

Is it the case that mechanisms which indiscriminately produced true beliefs, *at all times*, would be the *best* adaptive strategy? There is one sense in which the answer to this question is *yes*. Cases like Harry's do not show that true beliefs are sometimes *maladaptive*, rather, they show that not having *enough* true beliefs is maladaptive. If Harry had a true belief about the fate of the aeroplane, presumably, he would not have boarded it. Stich pre-empted this point by claiming that 'the question at hand is not whether omniscience would foster survival but whether more true beliefs are always better than fewer' (Stich 1990: 123). If that *were* the question, then of course an appeal to omniscience is no way out. However, if the question is: is selection of mechanisms which produce true beliefs the optimal adaptive strategy? Then, assuming omniscience, we can answer this question in the positive.

Putting aside biological omniscient creatures then, let us rephrase the question: would the selection of mechanisms which produce true beliefs be the optimal adaptive strategy *for us*? There are reasons to think that it would not. Stich gives us a case in which Harry has a true belief which turns out to be seriously maladaptive. Let us say that Harry has belief-producing

mechanisms which function to produce true beliefs. Is this the optimal adaptive strategy? It is worth looking briefly at Barry Loewer's objection to Stich on this point. He suggests that Harry's true belief that the plane leaves at 7:45 is similar to a gamble. The gamble is that there is a certain probability that holding this belief leads Harry to catching a plane and surviving, and a much smaller probability that this belief leads him to catch a plane in which he dies. Given that the probability of the latter state of affairs is much smaller than the probability of the former, the gamble 'has positive instrumental value *even if the belief leads to his taking a plane which crashes*' (Loewer 1993: 270). So having mechanisms which indiscriminately produce true beliefs is the optimal adaptive strategy insofar as truth seeking is the policy with the greatest expected utility (Loewer 1997: 277).

However, a little reflection reveals that this is not right. Recall Stich's idea of TRUE**** beliefs. In Harry's case, the optimal adaptive policy, *for him*, is to have mechanisms for belief-production which map beliefs to their TRUTH**** conditions. For Harry at least, having TRUE**** beliefs is more adaptive than having true ones (Stich 1990: 123), and would not require the 'gamble' as described by Loewer. Now of course, natural selection does not work at the level of individuals, it cannot be the case that given Harry's circumstances, mechanisms for TRUE**** beliefs are selected, whilst for Fred, in different circumstances, mechanisms for TRUE^^ beliefs are selected. However, what we can say is that, mechanisms for indiscriminate true belief-production is, quite clearly, not always the best biological strategy for an individual. **

Why might this be a problem for my account? Well, if my account were based on the claim that believing truly *all the time* was the most adaptive strategy for actual world believers, then it would, quite clearly, be based on a false claim. However, nothing I want to say about the nature of belief rests on this dubious claim. Indeed, it is consistent to claim that mechanisms for belief-production have the relational proper function of producing beliefs with true contents—these mechanisms are thus *supposed to* produce beliefs with true contents—whilst denying the claim that this is the most adaptive strategy. After all, it is false that natural selection 'can be relied upon to produce the best of all possible options, or even one that is close to the best' (Stich 1990: 97). Further:

[t]here are many factors in addition to natural selection that drive biological evolution [...] and each of these is capable of leading evolution away from an optimal phenotype. Moreover, even when natural selection is the only force at work, it cannot be counted upon to select the best option among those available. Nor can it be taken for granted that the best option *available* is the best *possible* option – or even that it is a particularly *good* option. (Stich 1990: 96-7)

There is another question which we ought to consider, that is, does natural selection, in fact, favour truth-directed belief formation? On my account the answer to this question is *yes*, natural selection does favour truth-directed belief formation, and it is from this claim that I derive my functional claim. But this is not to say that this is the optimal adaptive strategy, and indeed, my account does not require this claim. A better strategy would be TRUTH**** for Harry and TRUTH^^^ for Fred. Though, I agree with Loewer, that though the production of true belief might not be the *best* adaptive strategy, it is quite good:

although Stich correctly argues that natural selection cannot support the claim that our doxastic policies (the ones which have evolved) are optimal it does seem to support the claim that they are pretty good [...] if humans had typically found themselves in situations like Harry's they wouldn't have survived long enough to have invented airplanes. (Loewer 1993: 280)

I have shown in this section that selection of mechanisms which produce true beliefs is not the optimal adaptive strategy, but nor does it need to be for my account of belief to hold.

5.2.3 (F2) Our Mechanisms for Belief-production were Selected for Producing True Beliefs

I argued in §5.2.2 that true beliefs are adaptive. However, this is not enough to secure claim (F2), that our mechanisms for belief-production have in fact been selected for producing true beliefs. I move now then to a defence of this more specific claim. I need this claim as I appeal to our mechanisms for belief-production having the relational proper function of producing true beliefs, in virtue of their being selected for so doing, in order to give an explanation of the presence of Transparency and derivatively, an explanation of the Uncontrollability Thesis. I cannot get to the claim that our mechanisms for belief-production have been selected for the production of true beliefs merely by an appeal to true beliefs being adaptive, because something's merely being adaptive is far from a guarantee that it will be selected for.

Natural selection cannot guarantee the production of mechanisms which produce true beliefs, for such mechanisms may not have been available in the evolutionary past to be selected for (Sage 2004: 100). Therefore securing claim (F1), that these mechanisms were selected for, as well as the claim that true beliefs are adaptive, is not sufficient to secure claim (F2), that is, these considerations 'cannot guarantee that human cognitive faculties are truth-reliable' (Sage 2004: 110). Note though, that we do not need truth *reliable* cognitive faculties; it need only be the case that in certain key moments, it was the truth of beliefs which gave the creature a selective advantage. Recall, a Normal explanation for the performance of a function tells 'how it was

(typically) historically performed on those (*perhaps rare*) occasions when it was properly performed' (Millikan 1989b: 284, my emphasis).

Also note that something's being adaptive does not entail that it was selected for, it does not mean that there were even certain key moments in which it made a difference, I will call this the *No Guarantee Claim*. Because of this claim, we need an independent reason for thinking that natural selection has in fact selected for mechanisms of true belief production. We have said so far that true beliefs are adaptive (§5.2.2), and our mechanisms for belief-production have been selected for (§5.2.1). Though because of the No Guarantee Claim, this leaves room for the possibility that they were selected for producing false beliefs, or approximately true beliefs, or amusing beliefs, and so on. My task then, is to show that they were selected for the production of true beliefs. All I can do here is try to show that claim (F2) is more plausible than its alternatives, given that it is an empirical claim, and no amount of theorizing will prove it beyond doubt. Here then, I will argue that the claim that our mechanisms for belief-production have been selected for the production of true beliefs is the most plausible hypothesis.

Firstly we can once again consider the biological expensiveness of our belief-producing mechanisms. We can look to the relative cost-effectiveness of belief-producing mechanisms which produce largely true beliefs and lead to largely successful actions, as opposed to the cost-effectiveness of belief-producing mechanisms which produce largely false beliefs and lead to largely successful actions. Fales argues that the first kinds of mechanisms, those which produce largely true beliefs, are much more cost-effective than those which produce false beliefs because

there are no effective algorithms connecting false belief to appropriate action as there are when the input is true beliefs and the rules of inference employed are valid or inductively sound. Intelligent action is hard enough for a brain to manage; burdening it with ever-changing completely arbitrary principles would make the task impossible. (Fales 2002: 51)

Fales's claim here is that mechanisms which produce true beliefs, and successful action, are much more efficient and cost-effective than those which produce false beliefs, and successful action. This might be right, but it can only be our starting point. There are three issues with Fales's argument, each of which show that the argument is not sufficient to secure claim (F2). The first issue is Fales's assumption that our actions are largely successful. He claims that mechanisms which produce true beliefs would be a more economical way to get at successful action from the point of view of natural selection. Given that our actions *are* largely successful, so the argument goes, it is far more likely that our mechanisms for belief-production are geared to the production of true beliefs. For this line to work we need to be entitled to the claim that

our actions are largely successful. It is difficult to assess this claim, in part because we have no point of comparison. Nevertheless, the claim is a plausible one. If we consider how many thousands of actions we perform every day (eating, drinking, looking, grasping, pushing, pulling, typing, and so on) and compare this number to those which we consider *unsuccessful*, I submit, that that number will be a relatively small one (how many times a day do we fail to eat, drink, look, grasp, push, pull, type, and so on, when we attempt to?). I think we should grant to Fales the claim that our actions are largely successful.

Again, we should note that it need not be the case that our actions are *largely* successful, but only that the extent to which they are will make mechanisms aimed at producing true beliefs successful. Fales is working backwards from a stronger position—our actions *are* largely successful, so what best explains this? His answer is that our mechanisms for belief-production have been selected for the production of true beliefs.

Fales also assumes that natural selection would opt for the most efficient way of achieving an adaptive solution, but narrow scope efficiency is not the only criterion.⁵⁵ It might be more efficient for natural selection to select certain efficient wider mechanisms at the cost of less efficient mechanisms for belief-production. So Fales is not entitled to the claim that natural selection would opt for the most efficient way of achieving an adaptive solution. It could be true that our actions are largely successful, and still be the case that our mechanisms for belief-production are not truth-tracking. However, Fales does not need to discount this possibility here, as he does so a little later on, as we shall see.

Fales's third and gravest error is his comparison to falsity. It might be true that 'there are no effective algorithms connecting false belief to appropriate actions' (Fales 2002: 51), but that is not the only comparison available. Showing mechanisms geared to produce *true* beliefs would be more cost-effective than mechanisms geared to produce *false* beliefs only, at best, rules out one of a number of alternatives. The most we can take from Fales is the claim that our mechanisms for belief-production have not been selected for the production of false beliefs, but we have not yet arrived at the claim that they were selected for the production of true beliefs.

This is a difficult claim to argue for, made more so by the suggestion that the complexity of our mechanisms for belief-production supports the claims that natural selection has *not* favoured

55 A great example of natural selection selecting a highly inefficient mechanism is given by Richard Dawkins: '[i]n a person, the route taken by the recurrent laryngeal nerve represents a detour of perhaps several inches. But in a giraffe, it is beyond a joke – many feet beyond – taking a detour of perhaps 15 feet in a large adult! [...] On its downward journey, the nerve [...] passes within inches of the larynx, which is its final destination. Yet it proceeds down the whole length of the neck before turning around and going all the way back up again' (Dawkins 2009: 360–2).

mechanisms geared to the production of true beliefs. For example, James Sage claims that the biological costs of having truth-reliable cognitive faculties may just be too high:

[t]ruth-reliable cognitive faculties come at a high price: (i) the brain requires oxygen, calories, and cooling, (ii) calculating detailed inferences (even with minimal data) requires considerable time and concentration, (iii) accessing information from past experience requires extensive storage capacity and retrieval pathways, (iv) identifying relevant information requires multi-level sorting subroutines, (v) ranking desires and goals requires extensive deliberation and reflection, and (vi) utilizing “detectors” (and other perceptual inputs) requires precision and acuity. Each of these factors carries a significant biological cost. Since biological resources utilized by truth-reliable cognitive faculties could be used in other ways to increase inclusive fitness (ways that would confer an immediate benefit to the organism), it follows that natural selection may favor fitness-reliable cognitive faculties that are *not* truth-reliable. (Sage 2004: 104)

A similar claim comes from Stich, he claims that ‘strategies of inferences or inquiry that do a good job at generating truths and avoiding falsehoods may be expensive in terms of time, effort, and cognitive hardware’ (Stich 1990: 61). It might be that less truth-reliable systems which ‘do a less good, but still acceptable, job of generating truths’ might be preferred by natural selection (Stich 1990: 61).

However, earlier I noted that it need not be the case that our mechanisms for belief-production are truth *reliable*, but only that at certain key moments, their producing true beliefs was adaptive, and so it might not be that we need all the resources Sage and Stich point to. It need only be the case that a small set of beliefs (say those about food, sex, warmth) would be adaptive (recall that it seems to be part of the evolutionary design of our mechanisms for belief-production that ‘our dispositional beliefs come in varying strengths’ (Millikan 1995b: 251)). So the resources Sage and Stich point to may not all be in play in cases in which the costs of error are low. Next I will argue from the bigger claim that our mechanisms for belief-production are largely truth-reliable, in order to get to the smaller claim that their being truth-reliable for just a small set of beliefs such that they were selected in order to be so, is also true.

Fitness reliable mechanisms which are not truth-reliable are highly implausible. Fales offers a challenge to those who claim that mechanisms selected for the production of false beliefs could be equally adaptive as mechanisms selected for the production of true beliefs. The challenge is to construct a detailed account of how a creature (Fales uses a monkey), presented with a food source, forms a false belief which meets the following conditions:

- 1) combined with other present false beliefs and/or destructive desires, it leads (with good probability) to felicitous action.
- 2) when combined on other occasions with yet *other* false beliefs/bad desires, it *still* is likely to produce correct action; and
- 3) if destructive desires are invoked, a plausible Darwinian story can be told about how they evolved from the action-guiding desires of the pre-rational ancestors of the monkey, or in some other way. (Fales 2002: 52)

Once this is achieved, the opponent must then ‘construct a system of algorithms that will achieve this for the monkey’s beliefs generally’, Fales claims that this ‘can’t be done’ (Fales 2002: 52). Fales is not denying that sometimes false belief can lead to felicitous action. Rather, the point is that if false beliefs do some useful work, they do not do so ‘*in accordance with any Normal explanation*’ (Millikan 1995d: 72). Rather, a belief’s being true is ‘one of the conditions that must obtain if the belief is to fulfil any such functions in accordance with a Normal explanation’ (Millikan 1995d: 73). That is, ‘when actions out of false beliefs are successful, that’s generally a *lucky accident*’, and, correspondingly, that a policy of acting on false beliefs, even when it works in the short run, generally gets you into trouble sooner or later’ (Fodor 2002: 38). We saw a similar claim from Dokic and Engel in chapter one, in their discussion of Ramsey:

a false belief can only be useful locally and by accident. A false belief cannot properly guide our behaviour in every possible situation. In contrast, truth systematically promotes the success of action, on a large scale. Truth offers a *guarantee* of success that falsity cannot hope to deliver. (Dokic and Engel 2002: 48)

Again, we must be careful not to compare the production of true beliefs with the production of false beliefs, as this is not the only alternative. The question then must concern ‘whether or not a process of natural selection is likely to generate cognitive mechanisms that are systematically unreliable but somehow prove adaptive’ (Ramsey 2002: 20). Could we get to an account of ‘how an unreliable cognitive system can be *systematically* adaptive?’ (Ramsey 2002: 21) These kinds of mechanisms are certainly not impossible, but nor is an appeal to them a good explanation of the adaptiveness of our behaviour.

What we can say then is that it is far more plausible, for reasons we have seen, that natural selection has selected for the production of true beliefs, because our behaviour is largely successful, and an account of how unreliable mechanisms for belief-production could nevertheless produce largely successful behaviour is hard to come by. The best explanation is

that our mechanisms for belief-production are truth tracking ones.

My claim then is that creatures with mechanisms for belief-production which produce beliefs with true contents have proliferated because—given the essential nature of belief (its motivational role)—creatures with true beliefs will be more likely to survive because their actions will be successful. In Millikan’s terminology, I claim that the mechanisms in us which produce beliefs have, as one of their proper functions, the *relational proper function* of producing beliefs with true contents.⁵⁶ Other ways of characterising the function of our belief-producing mechanisms would suggest that such mechanisms would not have the same selective advantages across the board as they do in fact have. Variations on how well the function is performed can be accommodated by the differential setting of confidence levels regarding what is required for something to be taken as true.

Before I can fill in the explanatory work my account can do, it is worth pre-empting an objection which arises out of the possibility of the truth of semantic epiphenomenalism. On this view, it could be true that beliefs have efficacy in relation to behaviour, but that they do not do so in virtue of their *content*. If the semantic contents of beliefs are epiphenomenal, they ‘will be invisible to natural selection’ (Plantinga 2002: 6–7). William Ramsey proposes an argument from semantic epiphenomenalism (which he later attempts to refute) which makes the threat to my account clear: what matters for the production of behaviour is what goes on inside a creature’s head, so if beliefs are efficacious in the production of behaviour, they will be so in virtue of their physical properties. Semantic properties are not reducible to such physical properties and so the truth values of beliefs will not be reducible to them either. It follows that for any given belief, its truth value is irrelevant to its causal role in the production of (adaptive) behaviour. Given that what matters for natural selection is adaptive behaviour, truth is a property of beliefs which makes no contribution to the creature’s survival or its mechanisms for belief-production being selected (Ramsey 2002: 17).

Plantinga, in pushing this line, credits Cummins with the claim that with regard to the relationship between belief and behaviour, semantic epiphenomenalism is the ‘received view’ (Plantinga 2002: 10). The motivation for this view is claimed to be that materialism makes it very difficult for the content of a belief to play a causal role in behaviour. If beliefs are identified with

56 For the sake of completeness: our belief-producing mechanisms also have adapted proper functions (as is the case with all relational proper functions), at any given time of functional performance. So, Normally, when these mechanisms produce a true belief, in so doing they perform their relational proper function (which is invariant) and they perform their adapted proper function (which is variant). As I am interested in explaining the behaviour of these mechanism in general—as opposed to any particular instance of functional performance—I will make little use of the idea of an adapted proper function.

neural structures, the contribution to behaviour of those structures would be the same even if the contents changed (Plantinga 2002: 10).

My claim is that our mechanisms for belief-production have been selected for producing beliefs with true contents. If semantic epiphenomenalism holds—so the argument goes—this is not a coherent story. Natural selection will be blind to the semantic properties of beliefs, and so will not be able to act upon them. My account would have to be revised then to say something like the following: our mechanisms for belief-production have been selected to produce beliefs with *certain syntactic properties*. This might not be thought to be a very illuminating claim, nor will it help in an explanation of the features of belief which indicate a link between it and truth.

I can respond by pointing out that the claim that true beliefs are adaptive does not mean that the *content* should be efficacious. It could be that ‘true’ picks up on syntactic properties (which natural selection is not blind to), which pick up on something in the environment. The claim that animals with true beliefs are more likely to survive does not involve commitment to the claim that it is the content of the beliefs that is efficacious. Rather, when beliefs are true in certain specified circumstances, the selective mechanism will be at work.

Given the norm of performance bestowed on items with proper functions (in virtue of their meeting my conditions (a) and (b) (§4.5)), our mechanisms for belief-production are *supposed* to produce true beliefs, insofar as the biological history of those mechanisms lays down a standard from which produced adapted devices (beliefs) can deviate. I think such mechanisms are members of a higher-order reproductively established family, produced by certain genes whose function it is to produce mechanisms with the relational proper function of producing true beliefs. Given the biological advantages enjoyed by a creature with mechanisms which produce true beliefs, the mechanisms’ producing such items, causally explains why we have these mechanisms now. Creatures with mechanisms for belief-production which produce beliefs with true contents have proliferated because—given the essential nature of belief (its motivational role)—creatures with true beliefs will be more likely to survive because their actions will be successful.

In the case of empirical beliefs, true beliefs seem to coincide with some kind of *matching relation* to the environment, so empirical beliefs might hold correspondence with the environment. I think that this is not too contentious, and accords reasonably well with intuition, Robert Stalnaker for one thinks that it does. He claims that it is ‘intuitively clear [...] that there is a presumption that people’s beliefs will correlate with, and be caused by, their environments’ (Stalnaker 1987: 19). Also, this way of understanding *true* belief tallies well with Millikan’s notion of a *relational proper function*. However, not all beliefs will hold this matching relation, and so I will

not commit to a correspondence theory of truth. Non-empirical beliefs can still be understood as adapted devices, even if it is not a *matching* relation they hold with the environment.

As we have seen, in order for a trait to have a relational proper function, its function must be the production of something which stands in a relation to something else (Millikan 1984: 39). Given the function I have ascribed to our belief-producing mechanisms, they thereby meet this constraint. The thing they Normally produce is beliefs with true contents. The Normal explanation for the performance of this function is that the creature acts on the true belief in such a way as to satisfy its desires (it is because beliefs are motivational states that their being true is biologically useful).

What are we to say of the beliefs which result from the mechanisms which produce them? Do they have proper functions? No—they are not reproductions. One might have a belief that *there is a pineapple under the car next to the telephone*—such a belief might be the first of its kind and so would not be a reproduction of a past belief. But beliefs are still *supposed to do* something—if they were not it would not be appropriate to apply a standard of correctness to them. What then, are they supposed to do? They are supposed to be true. Using Millikan’s terminology, I claim that the beliefs produced in us have the (invariant) *derived proper function* of having true contents.

Let us see how these function claims play out then, by looking at a very simple case. Fred is hungry—he has a desire for food. In order to satisfy this desire he needs to have some beliefs, more specifically, he needs to have some beliefs about where there is food. Fred sees some food to his left. Here, his belief-producing mechanisms perform their relational proper function of producing beliefs with true contents by producing the belief *there is food to my left*. The adapted proper function of Fred’s belief-producing mechanisms here is to produce a belief with the content *there is food to my left*. The adaptor in this case, which Fred’s belief-producing mechanisms produce something (a belief) in relation to, is Fred’s environment (which includes there being food to his left). Fred then comes to have the belief that *there is food to [his] left*. This belief has the invariant derived proper function of having true contents. All beliefs of Fred’s produced in line with this function have this derived proper function, in virtue of their being produced by mechanisms with the relational proper function of producing true beliefs. Adapted derived proper functions are those derived from the producer thereof plus the adaptor for that producer (Millikan 1984: 41–2). Hence Fred’s belief in this case also has an adapted derived proper function of having the content *there is food to my left*.

5.3 Function: Explanatory Work

In this section I will discuss the explanatory work that my function ascription can do. I argued earlier that teleological and normative accounts of belief cannot give an explanation of the three features of belief which indicate a link between belief and truth (chapters two and three). I have also argued that the motivational account of belief can be defended from objections, but that it alone cannot account for those features either (chapter one). How then does my account explain these features? These features were Transparency, the Uncontrollability Thesis, and Epistemic Normativity (though I noted at the start of the chapter that the work on Epistemic Normativity is to be done in chapter six).

I have said that our belief-producing mechanisms have a norm of performance bestowed on them in virtue of the causal-historical relationship they hold to previous mechanisms which met the three clauses in condition (b) of my proposal for functional kind membership (§4.5). They have the relational proper function of producing true beliefs. When mechanisms produce devices in accordance with their relational proper function, they produce devices which are *adapted to* something else. Our mechanisms for belief-production are *supposed to* produce true beliefs. They Normally produce devices which are *adapted to* the way things are in the creature's environment, that is, they Normally produce beliefs with true contents.

This account is going to take care of many of our non-deliberative beliefs, those whose content we have not deliberated upon, those we have in virtue of perceptual experience or inferential patterns of reasoning. The mechanisms in these cases produce adapted devices which play the motivational role of belief, without requiring deliberation from the agent. For example, when I walk into an unfamiliar room, I may form beliefs about whether or not there are other people in the room, whether or not there is furniture in the room, whether or not there are wild animals or threats to my life present, but I need not *deliberate* in order to come to hold these beliefs. Rather, these beliefs are produced non-deliberatively, with the mechanisms for such production producing devices which are *adapted to* the way things are, that is, that there are four people, three tables and no animals or threats in the room. This is not to say, of course, that we have an infinite set of non-deliberative beliefs. For example, I do not claim that when I walk into an unfamiliar room my mechanisms for belief-production produce in me beliefs that *there is not a dragon in the room, there is not a serial killer in the room, there is not a trap door in the room*, and so on. Here we can appeal to the distinction between *explicit* belief and *tacit* belief, this can be understood as the distinction between beliefs whose content is *explicitly represented* as opposed to beliefs whose content is a consequence of those explicit representations, and are not themselves explicitly represented. So the consequences of the belief that *there are no dangers in the room*, can

generate an infinite set of tacit, non-explicitly represented beliefs: *there is not a dragon in the room, there is not a serial killer in the room, there is not a trap door in the room*, and so on. These beliefs may well be avowed when the subject is appropriately cued, which will make them *explicit* beliefs whose contents are explicitly represented. It should also be noted that beliefs formed from perceptual experiences or inferences may be deliberative. For example, one considers the question *whether to believe that p* and one looks to see *whether p is true*, cases like these are deliberative beliefs settled by the question *whether p is true*, perception is the way of answering the question *whether p is true*. But some non-deliberative tacit beliefs are also perceptual.

In this regard, our mechanisms for non-deliberative belief-production behave in a similar way to the mechanisms for changing skin pigmentation in a chameleon. This is not to say that chameleon skin pattern and beliefs are directly analogous, I only use the analogy to demonstrate my line of argument. Again, I say nothing about the biological validity of this example (see fn. 47). Both have relational proper functions. The former has the relational proper function of producing beliefs with true contents, and the latter has the relational proper function of producing skin patterns which match the chameleon's environment. In both cases these functions are carried out in the absence of deliberation by the agent. A chameleon does not have to deliberate about the colour of the leaf it sits on in order for the skin changing mechanisms to produce an appropriate device, and we do not have to deliberate on whether there are other people in the room in order for our belief-producing mechanisms to produce a belief with that content. We have seen how my account can explain the truth-directedness of our non-deliberative beliefs. I now turn to deliberative beliefs, which are characterized by Transparency.

5.3.1 Transparency

My account of belief is a motivational one, claiming that beliefs have a certain motivational role, as defended in chapter one (§1.2). We saw earlier that a motivational theorist might take this claim as the starting position for seeking to explain why our deliberative belief formation is transparent to considerations of truth. She might claim that transparency to truth considerations in our doxastic deliberation might be a good way to get true beliefs, which are the kind of beliefs which are useful given the motivational role specified for them. Alternatively, she might claim that given my desire to act successfully, I focus on truth in my deliberative belief formation, recognising that having true beliefs might make my actions more likely to succeed. I argued that these strategies were not able to provide an explanation of Transparency (§1.4). However, with my function claim made, an explanation is now available.

The explanation I offer appeals to certain causal facts about belief formation, which obtain in virtue of natural selection selecting for mechanisms which produce beliefs with true contents. As a proponent of the etiological account of biological function, I claim that the fact that a device has a certain function is an explanation for that device's presence today. There are two related questions to ask about the presence of the Transparency in our doxastic deliberation. These two questions will demand different kinds of answer, in line with Dretske's distinction between triggering and structuring causes.

Transparency can be understood as a disposition to allow certain considerations to settle whether or not one has a belief. When a subject has sufficient evidence for p , this disposes her to arrive at a belief that p , so evidence for p generates a belief that p via Transparency. We can ask a *how* question regarding Transparency: how is there Transparency in any case of doxastic deliberation? This question demands an answer providing details of the *triggering* cause of Transparency, that is, an answer which points to the mechanisms responsible for causing Transparency, understood as the disposition to allow certain considerations to settle whether or not one has a belief. The explanatory burden here is to explain why, when a subject deliberates over *whether to believe that p* , she moves immediately to considering the question *whether p is true*. The triggering cause of Transparency—of this disposition—can be given by appeal to certain neurological structures which realise the disposition to allow certain considerations to settle whether or not one has a belief. In functional terms: Transparency is achieved by certain causal facts which hold for our mechanisms for belief-production, which make it such that whenever our mechanisms for belief-production Normally perform their relational proper function and produce a true deliberative belief, the adaptor in each case is the way the environment is, that is, epistemic considerations pertaining to the truth of p . All token instances of being disposed to allow epistemic considerations to settle whether to believe are instances of believers immediately and inescapably attending to the adaptor (the environment) for the adapted device (the belief).**

Note that this kind of explanation—an explanation of how Transparency is realised—is the same kind of explanation given by the teleological and normative accounts discussed in chapters two and three. All three explanations appeal to something which triggers our being disposed to move from the question *whether to believe that p* to the question *whether p is true*. On the teleological account, this disposition is explained by reference to the aim₁ one adopts in posing the deliberative question, on the normative account, it is explained by reference to the norm governing belief. I argued though that these explanations offered by the teleological and normative accounts do not work (§2.2, §3.4.3).

We can also ask a *why* question, that is, why are we disposed to move from the question *whether to believe that p* to the question *whether p is true*? What explains why the causal facts obtain, which realise this disposition? That is, *why* does Transparency characterize our doxastic deliberation, why are the neurological structures which realise it instantiated? This question demands an answer citing the *structuring* cause of Transparency, an answer which points to why Transparency characterises our deliberation and not something else, realised by a different set of neurological structures. This explanation will explain why certain neurological structures are present.

The structuring cause of Transparency can be given by appeal to natural selection, which selected for the particular neurological structures which realise the disposition to move from the question *whether to believe that p* to the question *whether p is true*. My account then, unlike the teleological and normative accounts, gives an explanation of Transparency which cites a structuring cause. Given that our mechanisms for belief-production have the relational proper function of producing devices with true contents, the story for their selection includes their Normally producing true beliefs. Why do they have this function? Because devices with relational proper functions produce devices which are *adapted to* something in the environment. In the deliberative case our cognitive architecture is arranged such that when we seek to deliberate as to *whether to believe that p*, we are only sensitive to the adaptor (our environment) because this makes the adapted device (the resulting belief) more likely to perform its derived proper function of being true. The neurological structures which secure Transparency have been selected for their role in producing true beliefs in the deliberative case.

As we have seen, an explanation of Transparency may be able to be given by appeal to the triggering cause in each instance of deliberation—on my account, the triggering cause is given by causal facts regarding neurological structures. However, as Dretske stresses, this is not the only kind of explanation available. My account can also give an explanation in terms of the structuring cause of Transparency, that is, what made it the case that these causal facts obtain. My explanation of Transparency, in terms of an appeal to the biological function of our belief-producing mechanisms, offers a structuring cause for the phenomenon. By appealing to this function, we can also give an account of the triggering cause of Transparency, that is, we can give an account of why certain causal facts about our belief formation hold. This is good explanatory work, Dretske claims that:

[f]unctional explanations, the sort we get from evolutionary biology, are surely consistent with the more proximal explanations of neurophysiology. Both can be correct, and both reveal part of the

truth. They do not compete with but complement one another. They merely deal with different sets of causally relevant factors. (Dretske 2004: 167)

My appeal to biological function explains the behaviour of our belief-producing mechanisms not by looking '*inside* for the physical cause of external change but *outside* for the events that shaped that internal structure, that made the system into what it is today' (Dretske 2004: 174). This contrasts with the kind of explanation teleological accounts and normative accounts are offering, given that they only offer explanations which cite a triggering cause of Transparency, explanations which do not work (§2.2, §3.3). In contrast, my claim is that we have a pre-commitment to truth when we deliberate, this pre-commitment is secured by causal facts which constitute the triggering cause of Transparency, of which I am also able to offer a structuring explanation.

5.3.2 The Uncontrollability Thesis

Now consider the Uncontrollability Thesis, a characterisation of our inability to believe at will. We saw in chapter one that this inability follows from Transparency holding, though it does not require it (in §3.3 I showed that the normativist can give explanations of the Uncontrollability Thesis though not Transparency, and so the former could be understood independently of the latter). So the Uncontrollability *follows from* Transparency, but does not depend on it, which is why I have treated these phenomena separately. The truth of the Uncontrollability Thesis falls straightforwardly out of our deliberative belief formation being governed by Transparency. If all our deliberative beliefs are governed by Transparency, it follows that one cannot have a deliberative belief not governed by Transparency—that is, that one cannot believe at will.

If beliefs are adapted devices, then to produce a belief by an effort of the will alone would be to produce a belief irrespective of [what would have been] its adaptor (i.e. the creature's environment). But if that is the case, it could not have been produced by our mechanisms for belief-production, because such mechanisms are *supposed to produce adapted* devices. Producing devices which are not adapted is precisely what they are not supposed to do. So producing beliefs at will is going to be amongst the things that our belief-producing mechanisms are not supposed to do. If the mechanisms in me which produce my beliefs have the relational proper function of producing *true* beliefs, they would fail to perform that function fairly systematically if I had and used the ability to believe at will, as they would produce states which do not hold an appropriate relation to my environment. This is not yet a problem, we know that functional traits malfunction. But we also know that in order to ascribe to a trait a proper function, it needs

to have originated as a reproduction of a prior trait, which too served that function, which contributed to the biological success of the bearer of that trait. Normal conditions for the performance of relational proper functions is not going to include the willed production of the adapted devices. In fact it is going to *exclude* it. It is a necessary condition on having a relational proper function that the mechanism in question produces a device which is *adapted to* something else—mechanisms which allowed willed for belief would fail to meet this condition. And it is this exclusion (among others) which will be part of the story for their selection.

As an aside, it is important to point out that I have not said anything about the biological usefulness of willed beliefs. All I have done is argue that given the sort of things beliefs (contingently) are, their being willed is impossible. This impossibility is still a contingent one, because beliefs being adapted devices is a contingent feature—one which holds for beliefs in the actual world, but not necessarily for beliefs across worlds. Even if it could be shown that the ability to will a belief would be a maladaptive ability (and I am not convinced that it would), it would be unwise to argue on this basis—something's being maladaptive is far from a guarantee that it will not occur. To be clear: in giving an account of the Uncontrollability Thesis I have said nothing about the possible maladaptiveness of believing at will. Rather I have appealed only to beliefs being adapted devices. In doing so it is legitimate to give an explanation of the Uncontrollability Thesis by appeal to the biological function of our belief-producing mechanisms, and thus understand the thesis as expressing a contingent claim.

5.4 Objections

In this section I will outline and respond to two objections to my explanations of Transparency and the Uncontrollability Thesis.

5.4.1 Explanatory Redundancy

The following objection might be posed for my explanation of Transparency and the Uncontrollability Thesis by appeal to biological function. My explanation appeals to certain causal facts about belief formation, with the reason for these facts holding to be found by appeal to the process of natural selection.

However, we can imagine a non-biological creature, with whom we do not share a biological history, for whom these same causal facts hold. If we want an explanation of why this creature's beliefs are governed by Transparency, for why this creature cannot believe at will, and so on, our explanation will be in terms of these causal facts. Given this, it looks like my appeal to evolutionary considerations cannot play a role in an explanation of the two features of belief

which indicate a link between belief and truth. They are just one realisation of a certain kind of explanation. My explanation is explanatorily redundant. (Note also that had the explanations for Transparency offered by the teleological and normative accounts worked, they would apply equally to this non-biological creature.)

The charge of explanatory redundancy is right, to a point. However, as we have learned from Dretske there are two kinds of causes, and thus two kinds of explanation. I can respond to the charge of explanatory redundancy with the claim that although appeal to the relevant causal facts *does* constitute one kind of causal explanation, an appeal to natural selection constitutes another kind of causal explanation. So though the causal facts can do some explanatory work insofar as they can explain what realises the phenomena, they cannot explain why the phenomena are present in the structuring sense. So in accounting for belief's relationship to truth, they cannot do all the work, whereas an account which has the causal facts and an appeal to natural selection, can.

How does this help in responding to the charge of explanatory redundancy? As I have said, it is right that the relevant causal facts in each case can explain Transparency and the Uncontrollability Thesis, but that is not to say that an explanation appealing to selective mechanisms is redundant. Rather, the latter constitutes a different kind of explanation to the former.

When we want to know what causes the disposition to move from the question *whether to believe that p* to the question *whether p is true*, we want to know what the triggering cause of this disposition is. In answer to this we can appeal to certain neurological facts, which cause the disposition. However, what if we want to know what causes the neurological facts which realise Transparency, rather than there being some other set of neurological facts which did not realise Transparency? That is, *why* does Transparency characterize our deliberative belief formation? Answering these questions will require an appeal to the structuring cause. We can appeal to natural selection which has selected neurological structures which realise the disposition to move from *whether p* to belief that *p*. These are different kinds of explanation, having different relationships to the effect:

[t]he structuring causal relationship is a one-many relation, whereas the triggering causal relationship is one-one. [...] With triggering causes, distinct effects are produced by distinct causes, and distinct causes produce distinct effects. (Dretske 2004: 170)

The triggering cause of Transparency is going to explain why a subject is disposed to move from the question *whether to believe that p* to the question *whether p is true*, at any given time. This causal

explanation is available to she who wants to explain the features of belief just by appeal to causal facts. However, we still need an explanation for why that disposition is instantiated. Causal facts relating to neurological structures are not going to offer this sort of explanation. What will offer it is an appeal to causal facts about selective mechanisms. Thus the charge that my account, in its appeal to natural selection is explanatorily redundant, is unfounded. It is not that evolutionary considerations are just one realisation of a certain kind of explanation, rather, they constitute a different and important causal explanation altogether.

This response leaves it open that there might be other structuring explanations, and hence that the link between belief and truth is not *essentially* to do with biological function, for example, God might make it the case that belief is transparent to truth considerations. He might ensure that we are disposed to move from the question *whether to believe that p* to the question *whether p*, by ensuring certain neurological structures are instantiated. Leaving this possibility open is all for the good: the connection to truth is not necessary, nor is the way in which it is contingently realised the only way that it might be realised. I leave it open that other world beliefs might too be connected to truth, that they could enjoy the three features of belief, without this being explained by biological history. The existence of these features as exhibited by actual world beliefs, though, is explainable by appeal to the biological functions of the mechanisms which produce them.

I could claim that it is essential to the nature of belief that beliefs are produced by mechanisms with the biological function of producing true beliefs, this is Price's position, and we might call it the *strong function account of belief*. Price argues that the functional claim about belief—that our mechanisms for belief-production have the function of producing true beliefs—is a claim about the necessary nature of belief (Price manuscript: 9). This means that if an attitude is not produced by mechanisms with this biological function, whatever that attitude is, it is not a belief. Given this, there is no charge of explanatory redundancy for the strong function account of belief, because Price would not grant that there could be a non-biological creature with beliefs, and so for her there would not be an explanatory project concerning the features of those creatures' 'beliefs' which could rival the explanation offered by the functional approach to beliefs of biological creatures.

However, this alternative account is problematic for a different reason. Price would not grant to my objector that a non-biological creature could have beliefs, governed by Transparency or otherwise. However, presumably she would grant the metaphysical possibility of a creature for whom the same causal facts held. If such a creature were physically and functionally identical to a biological creature, one might think that this creature had beliefs.

Where my account does better than a stronger one like Price's then is that it can allow for this because I take the functional claim to be a contingent one. Providing the state in question could perform the appropriate motivational role (as defended in §1.2), that state is a belief.⁵⁷

5.4.2 Contingency is Unacceptable

I have argued that my account can explain Transparency and the Uncontrollability Thesis, by giving an explanation which appeals to natural selection as a structuring cause of the neurological structures which realise these phenomena. Given that I explain these features by appeal to biology, they come out as contingent features of the beliefs of some *believers* (i.e., those believers with a biological history), not as necessary features of *belief*. In this section I will consider an objection which states that this consequence of my account is unacceptable, and is a reason to reject it.

With respect to the Uncontrollability Thesis, the consequence of my account is that the thesis expresses a contingent claim about some believers, this is at odds with many philosophers' conception of our inability to believe at will. Indeed, '[m]ost' philosophers take it that 'our inability to bring about a belief just like that is a conceptual matter' (Scott-Kakures 1993: 77), and 'there is a widespread sense' that 'there is something in the nature of belief that makes it impossible to decide to believe a proposition for which one lacks epistemic support' (Frankish 2007: 528). More strongly: '[t]here is [...] something so chokingly unswallowable about the idea of someone's voluntarily coming to believe something that I have to suspect that this is ruled out at a deeper level than the contingent powers of our minds' (Bennett 1990: 90).

So one might resist my position on the grounds that it is a datum that Transparency and the Uncontrollability Thesis hold as a matter of metaphysical necessity, a datum which would ideally be explained, but equally, one which ought not to be abandoned because of failures to explain it.

There are a couple of things to say here in response, to justify my moderation of the necessity of the claims relating to Transparency and the Uncontrollability Thesis. Firstly, as I argued in chapters two and three, other accounts of belief have failed to offer good explanations of Transparency and the Uncontrollability Thesis (at any strength). Second, since the motivational account is right, as argued in chapter one, there is no motivation to add to it with

⁵⁷ In this respect, my account of belief is similar to Papineau's version of teleosemantics. Just as the 'essential core of teleosemantics' is a claim about states in the actual world (Papineau 2001: 286), so to is the biological part of my account of belief to be taken to be a claim about beliefs in the actual world.

respect to providing a demarcating criterion of belief. The motivational role specified for belief captures an important generality across creatures, with the link to truth as an add on, and so we must accept that Transparency and the Uncontrollability Thesis are contingent truths. If my opponent wants to insist that these phenomena reflect necessary truths about belief, then she can take both the motivational role of belief, as well as Transparency and the Uncontrollability Thesis, to be necessary features of belief. However, there is no motivation for doing this, and it would just be to project the particular circumstances of our biological heritage into a modal claim about belief. It would also rule out non-biological believers. So given that there are no good reasons to be committed to the necessity claim about these two phenomena, then an account like mine which casts them as contingent should not be objected to just on the basis of not honouring the necessity claim. If my account is able to do the explanatory work with respect to Transparency and the Uncontrollability Thesis, then the fact that it does not honour a thesis for which there are no good arguments is not a point against it.

5.5 Conclusion

In this chapter I claimed that our mechanisms for belief-production have the relational proper function of producing true beliefs. I defended this claim first by arguing that these mechanisms were selected for, and then by arguing that these mechanisms were selected for the production of true beliefs. I explained how this function ascription could give an explanation of Transparency which cites both the triggering and structuring cause of Transparency. I explained the Uncontrollability Thesis, which falls out of Transparency, by appeal to the fact that actual world beliefs are adapted devices. Finally, I defended my explanations from two objections. The first was that my account is explanatorily redundant, because it demonstrates just one way of realising a triggering cause. The second was that Transparency and the Uncontrollability Thesis coming out as contingent features of belief is an unacceptable result. I replied to both of these objections. In conclusion then, my account can give an explanation of Transparency and the Uncontrollability Thesis.

6

USEFUL BELIEF AND EPISTEMIC NORMATIVITY

In the previous chapter I argued that our belief-producing mechanisms have the relational proper function of producing true beliefs. So beliefs are (contingently) adapted devices with the derived proper function of having true contents. I then showed how my account could explain the presence of Transparency and the Uncontrollability Thesis which indicate a link between belief and truth.

In this chapter I will consider the objection that truth is not the only end for our mechanisms for belief-production, and so we need to recognise a second function: the function of producing *useful* beliefs (though not useful as an approximation to truth). I will distinguish between weak and strong versions of Epistemic Normativity, and claim that the function of producing true beliefs can explain Weak Epistemic Normativity. I will argue that (the illusion of) Strong Epistemic Normativity can be explained by appeal to the function of producing useful beliefs. I will conclude that my account can give an explanation of Epistemic Normativity.

6.1 Truth is Not the Only End

An objection to my account might be that the function of producing true beliefs cannot be the sole function of our belief-producing mechanisms. I will answer this objection by appealing to a second function proper to our mechanisms for belief-production: the function of producing *useful* beliefs. Note that when I refer to the function of producing useful beliefs, I do not mean useful *as* an approximation to truth. Rather, I will use ‘useful belief’ as an abbreviation for something like ‘self-directed effective functioning assisting belief’.

The claim that the only function of our mechanisms for belief-production is the production of true beliefs is vulnerable to many counterexamples. Of course, it is no counterexample to point to false beliefs, for claiming a function for a set of cognitive mechanisms says nothing about how often (or even if) this function will be performed. As Millikan puts it:

a description of the biological functions of the cognitive systems will in no way resemble a catalogue of psychological laws. It is certainly no psychological law, for example, that our beliefs are true, though it is a (teleo)function of our belief-fixing systems to fix true beliefs. (Millikan 1995a: 177)

So merely false beliefs do not constitute a problem for the claim that it is a relational proper function of our belief-producing mechanisms to produce true beliefs. Nor do widespread sets of false belief constitute a problem for this claim. This is because, as we know, functions can fail to be performed, and something possesses a function because *in certain key moments*, the performance of it contributed to the reproductive success of its bearers. (Compare with the case of sperm: it is no objection to the claim that sperm have the proper function of fertilizing ova that they rarely perform this function—‘[m]ost never find an ovum and have to call it quits’ (Millikan 1984: 34)).

Rather, what constitute counterexamples are those cases of false belief which seem to have been produced by mechanisms doing what they were *supposed to do*, that is, producing adaptive false beliefs. If there are cases in which our mechanisms for belief-production are functioning Normally when they produce false beliefs, this falsifies the claim that their *only* function is to produce true beliefs.⁵⁸ There are a huge number of examples of these kinds of cases.

Let us follow Allan Hazlett in calling a *doxastic practice* a ‘way of forming, sustaining, and revising one’s beliefs’ (Hazlett 2013: 41). We can think of these practices as ‘dispositions to form beliefs of given types in given natural circumstances’ (what Papineau calls a ‘belief-forming habit’ (Papineau 1987: 124)). We can call a practice a *doxastic bias* if it is an unreliable (in terms of truth) doxastic practice. If there are cases of non-truth tracking belief in which the mechanisms which produced them were doing what they were biologically supposed to do, that is, cases in which the mechanisms perform Normally, we can say that the mechanisms exhibit a doxastic bias in such cases. If there are cases of doxastic bias in which our mechanisms for belief-production are functioning Normally, it is false that the only function of our mechanisms for belief-production is the production of true beliefs.

An example of a widespread doxastic bias is self-enhancement bias, a term which includes ‘overly positive self-evaluation, unrealistic optimism, illusions of control, self-serving causal attributions, valence biases in recall and processing speeds, biased attention to evidence, [and] biased self-focused attention’ (Hazlett 2013: 52). For example, evidence of overly positive self-evaluation can be found in the oft-cited study by Lewinsohn et al (1980) on self-perceptions of depressed and non-depressed subjects, in which the self-ratings of participants were compared across various dimensions with the ratings of those same participants given by other people.

⁵⁸ Though I do not make the claim that the production of true beliefs is the only function of our mechanisms for belief-production (see fn. 49). I am only interested in the functions of these mechanisms which can explain the three features outlined in chapter one. I think we need to appeal to at least two functions in order to do this work.

Famously, what was found was that the ‘initial self-perceptions of the depressed subjects were less discrepant with observer ratings’ than the control (non-depressed) subjects (Lewinsohn et al 1980: 210). The depressed subjects’ self-ratings ‘did not differ significantly’ from those of their observers, whereas the control subjects rated themselves ‘significantly more positively’ than did their observers. The depressed subjects, then, were the ‘most realistic’ with regard to their self-perceptions, whereas the control subjects ‘were engaged in self-enhancing distortions’ (Lewinsohn et al 1980: 211). In his discussion of this empirical work, Peter Railton claims that ‘[i]t would appear to be part of the normal, healthy operation of one’s self image that one discount negative evidence and defy the odds’ (Railton 1994: 92). Several further studies support these findings, and have shown that most people consider themselves to be above average, or better than most other people, when asked about positive traits and abilities (Alicke 1985, Brown 1986, Dunning et al. 1989, Helgeson and Taylor 1993, Sedikides 1993, cited in Hazlett 2013: 44).

Similarly, people tend to be over optimistic about their futures, in particular, they underestimate the likelihood of experiencing negative events and overestimate the likelihood of experiencing positive events (Hazlett 2013: 47). Having inaccurate views about our abilities and life chances is a doxastic bias which may be adaptive. Papineau gives the example of the belief that ‘you are not going to be injured in some unavoidable and imminent trial of violence’, suggesting that natural selection may have ‘bequeathed us an innate disposition to form this belief, even in cases where it runs counter to the evidence, in order to ensure that we will not flinch in battle’ (Papineau 1991: 40).

We might also think about *partiality bias*, which is differential doxastic treatment with respect to one’s friends over strangers (Hazlett 2013: 89). Someone expresses partiality bias when she forms beliefs about her friend which are different from the beliefs she would form about a stranger, on the same evidence. Partiality bias may be a way in which our confidence thresholds for belief are affected (§2.3.1.2).

Finally, we can also consider self-deception, a capacity which is prevalent. There are broadly two ways of thinking about a biological explanation of self-deception. On the first way the capacity for self-deception evolved, because it is adaptive. On the second way the capacity for self-deception is a spandrel on other adaptive features. Robert Trivers (2000, 2011) gives a version of the first account, and van Leeuwen (2007b, 2008) a version of the second. Whatever account we prefer, we have a case of belief in which the mechanisms which produce them are performing Normally by producing (usually false) self-deceptive beliefs.

Clearly then, the mechanisms responsible for belief-production are not solely geared, in all cases, towards truth. My account of belief, insofar as this is the only function ascription it offers, fails to account for a large set of beliefs of ordinary people, which exhibit the kinds of biases discussed above. This is not yet a problem, my thesis is concerned with giving an explanation of the connection between belief and truth. If I can do this by appeal to the relational proper function of producing true beliefs had by our mechanisms for belief-production, then my account is complete *insofar* as it explains what it set out to. However, the function of producing true belief cannot explain the presence of Epistemic Normativity in a stronger sense (to be outlined below). The account then is incomplete. I have discussed cases in which our mechanisms for belief-production perform Normally in producing beliefs which do not track truth. This is because I think that the function that the mechanisms perform in these cases, is a function which can also be appealed to, alongside the function of true belief, to give an explanation of Epistemic Normativity.

The above considerations then show that we need to recognise another function proper to our belief-producing mechanisms. I suggest that it is the function of producing *useful* belief, understood as I said earlier as something like *self-directed effective functioning assisting belief*. This function is performed by our mechanisms for belief-production when beliefs like those cited above are produced. Let me now call the production of true beliefs discussed in the last chapter *proper function one*, and the production of useful beliefs to be discussed in this chapter *proper function two*.

Now let us consider self-enhancement bias and partiality bias. Hazlett suggests that just as there may be coping mechanisms in the form of self-deception to somewhat offset the negative consequences of bad life events, these mechanisms might also give rise to 'less extreme' biases which might be 'useful as [a] means of coping with the events of everyday life' (Hazlett 2013: 61). He suggests that everyday life is something to be *coped with* in so far as 'ordinary human existence is full of frustration and indignity, annoyance and insult. It *is* something with which we need to cope' (Hazlett 2013: 61). He cites work on self-enhancement bias carried out by Taylor and colleagues (2003), which found that 'self-enhancement bias was correlated with lower blood pressure, lower heart rate, and lower HPA (hypothalamic-pituitary-adrenocortical) levels, in response to stress', the stressful situation was the subjects having to count backwards in 7s from 9,095 (Hazlett 2013: 61). Such health effects are adaptive ones. With respect to partiality bias, it may be better for us to require higher standards of evidence when it comes to friends, rather than forego the relationship or have negative feelings towards friends. Again, my claim here is that beliefs based on self-enhancement bias and partiality bias are produced by our mechanisms

for belief-production producing beliefs in line with proper function two. The performance of this function occurs when circumstances are abnormal for the performance of function one. That is, they are circumstances to which we do not need to appeal when describing how function one was properly performed. Instead, circumstances in which truth can hurt the effective functioning of a subject are Normal circumstances for the performance of function two.

Let us look now to self-deception. The capacity for self-deception might be produced in line with the capacity for self-enhancement bias. This fits with van Leeuwen's account of self-deception as a spandrel on other features of mind. According to van Leeuwen, the first of these features is the 'sting' which accompanies desires; felt when anticipation of their non-fulfilment is present, which can help motivate the achieving of one's goals. The second is the ability to selectively attend to evidence, which is essential to any creature with finite cognitive resources and interests. Thirdly, humans have an inclination toward harm avoidance, the function of which is to keep us away from situations in which we may be harmed (van Leeuwen 2008: 199). van Leeuwen claims that these three features can work together to give rise to self-deceptive belief, for example: I desire that p and so consequently, upon anticipation that $\sim p$, I feel the sting (feature one). Such anticipation may be supported by evidence that $\sim p$. Discomfort is felt when I attend to the evidence that $\sim p$ but fades when I do not—particularly when attending to the evidence that p . Consequently I attend to the evidence that p and self-deception ensues (features two and three). I focus attention on the evidence that p and so come to believe that p despite the total evidence pointing to $\sim p$ (van Leeuwen 2008: 198–9). If this is the right account of the presence of the capacity for self-deception, the mechanisms which give rise to self-deception, do so in circumstances abnormal for the production of true beliefs, that is, circumstances abnormal for the production of beliefs in line with proper function one.

Alternatively, we might say that the capacity for self-deception is an adaptation rather than a spandrel. Trivers argues that there are a number of reasons why natural selection may have favoured mechanisms of self-deception. One such reason (and the reason which is given the most focus by Trivers) is that the capacity may make an organism a better inter-personal deceiver. An appeal to the mechanisms of ordinary deception is made, that is, deception between individuals, something which readily lends itself to an adaptationist account. Trivers claims that deception between individuals may generate self-deception, as a mechanism for preventing detection by others of ongoing deception (Trivers 2000: 115). The claim is that I can deceive my adversary better if I remain unaware that I am being deceptive, because if I am unaware, this leaves avenues for the detection of deception closed. On this approach then, the

conscious mind is construed as a social front, ‘maintained to deceive others who more readily attend to its manifestations than those of the actor’s unconscious mind’ (Trivers 2000: 115). Other reasons given for the capacity of self-deception to have evolved include reality distortion within the individual arising from parent-offspring conflict which may generate conflicting internal voices within a single organism (Trivers 2000: 115), and a self-deceptive positive stance towards one’s life being biologically advantageous.

Whichever account of the existence of the capacity for self-deception we prefer, my claim is that when self-deception occurs, conditions are abnormal for the production of true beliefs (in line with function one) and normal for the production of useful beliefs (in line with function two).

Though our mechanisms for belief-production have the production of useful beliefs as a proper function, that is not to say that we can recognise when forming such beliefs that that is what is going on. Our deliberation is characterised by transparency, so even if we deliberate over whether p , and come to self-deceptively believe that p , we will nevertheless take our deliberation to be answered by the question *whether p*. Recall that mechanisms with relational proper functions produce adapted devices, that is, devices which stand in some relation to the creature’s environment. I argued in chapter five that when our mechanisms for belief-production perform proper function one, this is a relational proper function, and the device produced (the belief) is an adapted device. Even when the mechanisms perform proper function two, though they are not producing adapted devices in the same way, doxastic deliberation is such that we do not distinguish first personally between beliefs produced in line with proper function one, and those produced in line with proper function two. There is a sense then in which we take all our beliefs to be adapted devices, we take them to, in some sense, correspond with, or reflect the environment—we take them all to be true.

Our deliberation being this way might be why beliefs produced in line with proper function two are not wildly inaccurate, inasmuch as they are not completely immune to evidence. Though self-enhancement bias might lead me to believe that I am an above average runner, it will not lead me to believe that I could compete in the next Olympics. Though I may require more evidence to believe that my friend stole from the food bank collection, I will not be able to continue to withhold belief in her guilt if I witness her stealing from the food bank collection, partiality bias ‘manifests itself only when you have inconclusive evidence, never when you have conclusive evidence’ (Hazlett 2013: 92). The same goes for self-deception, the cuckolded husband may self-deceptively believe that his wife is faithful, but if he is presented with

conclusive evidence that she is unfaithful, he will not be able to continue to believe that she is.⁵⁹ While the husband's desire for his wife to be faithful will be part of the causal story for his belief that she is, 'we would be hard-pressed to call his cognitions beliefs rather than fantasies if they were shown to be insusceptible to the evidence of his wife's transgressions when it were brought to his attention' (Shah 2003: 462). This is summed up with the following observation from Ziva Kunda:

[p]eople do not seem to be at liberty to conclude whatever they want to conclude merely because they want to. Rather, I propose that people motivated to arrive at a particular conclusion attempt to be rational and to construct a justification of their desired conclusion that would persuade a dispassionate observer. They draw the desired conclusion only if they can muster up the evidence necessary to support it [...] In other words, they maintain an "illusion of objectivity". (Kunda 1990: 482–3)

Deliberation being transparent to truth considerations, by the subject's lights, ensures that when mechanisms perform the second function proper to them, they only do so 'behind the scenes' as it were. So far as the subject is concerned, function one is being performed. I will now argue that we can explain the third feature of belief which indicates a link between it and truth—Epistemic Normativity—by appeal to both proper function one and proper function two.

6.2 Epistemic Normativity

In chapter one I outlined three features of belief which I suggested indicated a link between it and truth. In chapters two and three I looked at the teleological and normative accounts of belief, both of which contain the claim that the feature of belief they regard as essential to it (its being aimed or norm governed) could explain all three features. I argued that they could not.

In my alternative account put forward in chapter five, I claimed that our belief-producing mechanisms have the relational proper function of producing true beliefs, and I argued that their having this function explained why Transparency characterises our deliberative belief formation,

⁵⁹ One might think that subjects with Reverse Othello delusion—'a delusional belief in the fidelity of a romantic partner' (Butler 2000: 85)—are counterexamples to this point (and more generally, delusions might present a problematic case). Actually though, I do not think that they are. I think that delusions are beliefs, and they are beliefs on a continuum with ordinary beliefs, which do not differ in kind from ordinary beliefs in terms of their production and their rationality. It is beyond the scope of this thesis to argue for this claim, but if it is right, then subjects with delusions do not cause immediate trouble for the account developed in this thesis.

and derivatively, why the Uncontrollability Thesis holds. However, this leaves the third feature of belief unaccounted for.

Having recognised a second function proper to our belief-producing mechanisms, I am now in a position to answer this objection. Firstly, I do not think an explanation of Epistemic Normativity can be given solely by an appeal to the function of producing true beliefs, and it is this function which links belief to truth. The consequence of this is that Epistemic Normativity, as set out in §1.3.3, is not a feature of belief which links it to truth. Rather, it can be explained by appeal to *biological normativity*, and by appeal to the function of producing useful beliefs. In this section I will set out how this can be done.

6.2.1 Strong and Weak Epistemic Normativity

In chapter one, I said that the following two claims would be how I understand Epistemic Normativity:

(EN1) Beliefs have a standard of correctness.

(EN2) There are categorical epistemic norms.

There are at least two ways to understand claims (EN1) and (EN2), relating to how we understand the normativity involved. The target of my criticism is any position which takes claims (EN1) and (EN2) at face value, as involving *sui generis* irreducible epistemic normativity.

As we saw in chapter one, the standard of correctness in (EN1) is had by beliefs (and guesses, which I will put aside). I could have an imagining or supposing with a true or false content which is nevertheless not appropriately judged as correct or incorrect. I said in chapter one that with respect to (EN1) it is claimed that beliefs have a standard of correctness, that is, it 'is part of the "job description" of belief as a distinctive propositional attitude that beliefs are correct or incorrect depending upon the state of the world' (Railton 1994: 74). With respect to (EN2), the kinds of norms I have in mind include norms of rational belief formation, norms of justified belief formation, norms of evidence, and so on. Epistemic norms govern only belief—it is equally inappropriate to say of my imaginings or supposings that they are rational, irrational, justified, unjustified, and so on. These norms are thought to be categorical ones, ones which 'apply to agents categorically; their reason-giving force transcends agents' desires, ends, activities, or institutions', as opposed to immanent ones whose 'reason-giving force depends on agents' desires or ends, their engagement in certain rule-governed activities, or their occupation of certain roles' (Olson 2011: 80).

I consider Epistemic Normativity to place an explanatory burden on the belief theorist such that she needs to give an account of why (EN1) and (EN2) hold, or why we *think* that they do (when they do not). I will give an account which paves a middle way. There is a sense in which (EN1) and (EN2) are true, a sense which I will call *Weak Epistemic Normativity*, and this can be explained by appeal to proper function one. However, we *think* of (EN1) and (EN2) more in line with what I will call *Strong Epistemic Normativity*, and so as well as giving an account of why Weak Epistemic Normativity is true, I also need to explain why we make the mistake of thinking that Strong Epistemic Normativity is true. This mistake can be explained by appeal to proper function two.

Here then is my explanation of Weak Epistemic Normativity: biological norms lay down a standard from which token traits can deviate (§4.5). Our belief-producing mechanisms have the proper function of producing true beliefs, so there is a sense in which true beliefs are correct and false beliefs are incorrect, in the same way that there is a standard from which a heart can deviate which makes it a working or defective heart. Beliefs produced by mechanisms performing function one are adapted devices, the content of those devices can vary with respect to their meeting the correctness conditions laid down by biology. When a belief is true, something has gone right, mechanisms have performed their function, when a belief is false the mechanisms have failed to perform their function (or stronger, have malfunctioned). It is in this sense only that beliefs can be said to have a standard of correctness. This is an explanation of claim (EN1) of Weak Epistemic Normativity. We can also give a reductive account of the norms cited in (EN2). These norms are, in a weak sense, categorical norms insofar as they hold regardless of the interests of subjects. With respect to epistemic norms, my claim is that our beliefs about these norms, and our absence of beliefs about pragmatic norms for belief, are the result of biology. We reflectively endorse so-called epistemic norms, but we are mistaken about their strength. Insofar as we can make sense of norms applying to beliefs, there is no reason, on my account, why they could not be pragmatic (believe what makes you happy), given that belief is not constitutively or conceptually connected to truth. An appeal to biology explains why we endorse only what we take to be *sui generis* epistemic norms, because by doing so we are more likely to form beliefs which meet the standard of correctness laid down by biology. Our doing so makes it more likely that our mechanisms for belief-production will perform their relational proper function of producing true beliefs.

For want of better terms, I will call evaluative beliefs *about* beliefs *weak epistemic beliefs* (in line with Weak Epistemic Normativity) and *strong epistemic beliefs* (in line with Strong Epistemic Normativity). Epistemic beliefs are beliefs which express the claims of (EN1) and (EN2); beliefs

which have as their content a claim about the epistemic properties of beliefs or epistemic reasons for belief, such as ‘Glen’s belief that p is correct’ or ‘Tracey’s belief that p is unjustified’, or ‘Jean has an epistemic reason to believe that p ’. My claim is that strong epistemic beliefs are false, insofar as they are mistaken with respect to the strength of the normativity involved, this follows from my explanation of Weak Epistemic Normativity. I suggest that the etiology of strong epistemic beliefs provides an explanation for why we have them, even though they are false.

My account needs to explain Weak Epistemic Normativity, as I have done above, but it also needs to explain why we have the illusion of Strong Epistemic Normativity. There are immediate problems with trying to extend my biological account of belief to the illusion of strong Epistemic Normativity. Firstly, on the face of it, it is significantly trickier for me to give an account of the illusion of Strong Epistemic Normativity than it is for the normativist about belief to straightforwardly give an account of Strong Epistemic Normativity (not the illusion thereof), because the normativist takes claims (EN1) and (EN2) at face value, and gives an account of them at the strength of the normativity they imply. For the normativist, ‘normative properties of belief are constitutive of belief, and are thus explained by the very nature of belief’ (Chan 2013: 8). However, I deny that Strong Epistemic Normativity (and even Weak Epistemic Normativity) is constitutive of belief and believers. My claim is that there is no Strong Epistemic Normativity because it would follow that there would be norms which govern all believers. Rather, there are just weak epistemic norms which govern some believers, those with the right biological history. Strong epistemic norms then, make claims on all believers, weak epistemic norms only apply to believers with the right biological history. There is no reason to think that epistemic norms would apply to believers whose beliefs were not linked to truth, but we can make sense of the claim that they apply to believers whose beliefs are so linked, in our case, believers with the right biological history.

It might also look like the existence of the illusion of Strong Epistemic Normativity cannot be given a biological explanation given what I have said about Transparency, the Uncontrollability Thesis which falls out of Transparency, and Weak Epistemic Normativity. If a biological explanation of the illusion of Strong Epistemic Normativity involves giving an error theory grounded on the adaptiveness of strong epistemic beliefs, it looks like, given what I have already said, this kind of account is not available. With the causal facts which secure Transparency to truth considerations in place, what biological advantage would our having evaluative beliefs *about* beliefs have, at this particular strength? What biological advantage would our ascribing *sui generis* irreducible epistemic properties to beliefs have? From a biological point

of view, what difference would it make that we have false strong epistemic beliefs rather than true weak epistemic beliefs? Why would it be biologically useful to make a mistake regarding the strength of the normativity involved in belief?

My opponent might claim that if the question of *whether to believe that p* comes down to the question of *whether p is true*; that is, if Transparency holds, and if we had weak epistemic beliefs, strong epistemic beliefs have no adaptive role to play. If this is right then either Strong Epistemic Normativity (or the illusion thereof) does not pose an explanatory burden, or a biological account of belief excludes the possibility of an explanation for something which clearly requires one. The first possibility is an unhappy result and not one I would like to embrace—I have argued that Strong Epistemic Normativity places an explanatory burden on the belief theorist, either to account for it at face value (as the normativist does) or to explain a weaker version, and also the illusion of the stronger version (as with my preferred account). If a biological account of belief leads to the conclusion that there is no such explanatory burden, all the worse for that account. The worry for my account is that given what I have said already, I do not allow myself an explanation of the illusion of strong Epistemic Normativity, and thus my account cannot meet one of the explanatory burdens put to the belief theorist by belief's connection to truth (though as I have said, I do not think the illusion of Strong Epistemic Normativity is a feature which connects belief to truth, though it is a feature which requires an explanation).

I will argue that having strong epistemic beliefs is adaptive in ways not secured by the Transparency which characterises our doxastic deliberation. As such, I can give an Evolutionary Debunking Argument regarding our strong epistemic beliefs which complements what I have said about the second function proper to our belief-producing mechanisms. I will claim that this argument explains why we have false strong epistemic beliefs, and supports an error theory of these beliefs, according to which no strong epistemic claim is true.⁶⁰

6.3 Queerness

The motivation for my account of the illusion of Strong Epistemic Normativity will not rest entirely on the metaphysical queerness of normative properties, and hence the claim that we

⁶⁰ The work I am doing here is merely descriptive, I make no claims about what we ought to do with respect to belief's standard of correctness and the norms supposed to govern belief. I think a sensible naturalized epistemology would look something like the one Papineau develops according to which 'you should think of yourself as a system for generating true beliefs. You want to be as reliable such a system as possible. So you should consider ways of redesigning the system, and should implement those that promise an improvement' (Papineau 1987: 135).

cannot account for Strong Epistemic Normativity without invoking unacceptable queerness, but I here have an aside on this issue, which is worth bearing in mind for the discussion that follows. Considerations of queerness provide another reason for thinking that Strong Epistemic Normativity is an illusion. Below I give an Evolutionary Debunking Argument for strong epistemic beliefs which shows that such beliefs are unjustified. Considerations of queerness, as well as my account of Weak Epistemic Normativity, offer a further reason for thinking they are also false. My argument will be that a biological account can explain the illusion of Strong Epistemic Normativity without the cost of invoking *sui generis* irreducible normative properties, and that this is a theoretical advantage. If one is not convinced by arguments from queerness, at the very least, there is a theoretical burden on the proponent of Strong Epistemic Normativity which is not placed on my account, as I only allow Weak Epistemic Normativity. In this section, I outline this theoretical burden.

Hazlett discusses the problem of queerness, I will draw on his discussion of this widely talked about issue. I will follow him in taking a property to be *irreducible* if and only if it ‘cannot be reduced to natural properties’ (Hazlett 2013: 150), where we endorse a ‘relatively modest, if rough, conception of reduction: property P_1 is reducible to property P_2 when the instantiation of P_2 completely explains the instantiation of P_1 ’ (Hazlett 2013: 150). Hazlett suggests that as naturalists, we ‘have pro tanto reason to avoid positing irreducible normative properties’, he calls this the ‘[f]irst naturalistic constraint’ (Hazlett 2013: 150). He suggests that considerations of parsimony support the first naturalistic constraint—normative properties are not, at least not straightforwardly, part of the naturalist’s spatial-temporal system (Hazlett 2013: 151). Our reason to avoiding positing irreducible normative properties is only a *pro tanto* reason; an ‘ontology without rhinos might be simpler and more austere, but we’ve come face-to-face with too many rhinos for this to be plausible’ (Hazlett 2013: 151). Rather, there is a burden on the proponent of irreducible normative properties to justify this expansion of our ontology.

Hazlett also gives us a second naturalistic constraint, which is that ‘[w]e have pro tanto reason to avoid positing (causally efficacious) irreducible normative properties’ (Hazlett 2013: 152). This constraint is in place to avoid positing systematic causal overdetermination. But again, the reason we have is only a *pro tanto* reason; there is a burden on the proponent of efficacious irreducible normative properties to justify our allowing systematic overdetermination (Hazlett 2013: 152). My claim is that neither an expansion of our ontology, nor allowing systematic overdetermination can be justified, given that my account can explain Weak Epistemic Normativity *and* the systematic illusion of Strong Epistemic Normativity with an Evolutionary

Debunking Argument, and so there is no explanatory gain to be had from violating Hazlett's two naturalistic constraints.

If our strong epistemic beliefs were the products of our belief-producing mechanisms performing their first proper function—the production of true beliefs then the properties those beliefs cite would be instantiated in the world. This would be an unhappy result for the naturalist in light of worries about the metaphysical queerness of these kinds of properties; our epistemic discourse commits us to 'entities that are hard to square with a naturalistic world-view' (Olson 2011: 88). If my claim were that strong epistemic beliefs were produced in line with our mechanisms for belief-production performing their first proper function, I would violate both of Hazlett's naturalistic constraints. Regarding the first constraint, I would be positing irreducible normative properties, because beliefs like 'Mike's belief is justified' or 'Eileen's belief is rational' cite properties which are irreducibly normative. Regarding the second constraint, I would be allowing irreducibly normative efficacious properties, insofar as it is such properties I respond to when I form strong epistemic beliefs and make strong epistemic judgements.

We might also add that introducing irreducible normative properties is one thing, introducing *sui generis* normative properties is another. The proponent of Strong Epistemic Normativity does both of these things, they 'believe in sui generis norms of judgement, which prescribe the pursuit of truth independently of moral or personal considerations' (Papineau 1999: 2). We might ask the following questions with respect to this:

If these norms are quite distinct from moral or personal 'oughts', then where do they come from? What kind of fact is it that we categorically 'ought' to reason in certain ways? And whence does the guiding force of these 'oughts' derive—why *should* we reason in these ways? I know that I should do what is morally required. And there is an obvious sense in which I should do what will get me what I want. But I find myself in difficulty understanding why I should be moved by the non-naturalists' putative sui generis norms of judgement. (Papineau 1999: 27)

The proponent of Strong Epistemic Normativity then faces the usual worries raised by queerness, but also an additional problem in postulating a *sui generis* normativity which attaches only to belief.

Instead then, my claim is that strong epistemic beliefs are produced in line with the second

function proper to our mechanisms for belief-production, that of producing useful beliefs.⁶¹ Though again, the beliefs produced by these mechanisms are not completely wide of the mark, they are still truth tracking in some sense. Since the claims which constitute Weak Epistemic Normativity are true—biology lays down standards from which beliefs can deviate, in that sense, the mechanisms producing strong epistemic beliefs are getting something right. The mistake is in believing that Strong Epistemic Normativity is true. In the next section I will give an Evolutionary Debunking argument for strong epistemic beliefs. The primary purpose of giving this argument is to give an explanation of why we have strong epistemic beliefs, *even though they are false*. The argument will explain why we have these beliefs, even though it is only their weaker counterparts which are truth tracking.

6.4 An Evolutionary Debunking Argument for Strong Epistemic Beliefs

An Evolutionary Debunking Argument is an argument which seeks to undermine the justification of a set of evaluative beliefs by appealing to their evolutionary origins (Kahane 2011: 103). It will cite a non-truth-tracking process for the production of those beliefs which counts as an undermining defeater for them (Kahane 2011: 105–6). Guy Kahane offers what he calls the ‘basic structure’ of Evolutionary Debunking Arguments as:

Causal premise. S’s belief that p is explained by x.

Epistemic premise. X is an off-track process.

Therefore

S’s belief that p is unjustified. (Kahane 2011: 106)

I will fill in this structure in order to argue that we can give a naturalistic account of the illusion of Strong Epistemic Normativity with an error theory grounded on the off-track processes which result in our strong epistemic beliefs. As I said above, my primary aim in offering this kind of argument, is not to argue that the target beliefs are *unjustified*, it is rather to explain why we have such beliefs, even though they are false. However, the argument will also show that strong epistemic beliefs are unjustified. There might be a question of why this is necessary, given

⁶¹ Perhaps epistemic beliefs could be the result of mechanisms malfunctioning, or simply failing to perform disjunctive function one. However, I think that epistemic beliefs (and other beliefs produced in line with proper function two) are being produced Normally, that is, they are not malfunctioning devices, they are produced by mechanisms functioning Normally.

my claim that weak epistemic beliefs are true (it follows from this that strong epistemic beliefs are false), and considerations of queerness which indicate that strong epistemic beliefs are *false*. However, if strong epistemic beliefs were justified, this might provide traction against my claim that only Weak Epistemic Normativity holds. If we are justified in having strong epistemic beliefs, then perhaps we ought to look again at whether they are false, our being justified in having such beliefs might suggest that they are in fact true. Adding the claim that strong epistemic beliefs are unjustified then (where ‘unjustified’ is understood in a purely descriptive way), adds weight to my account. Now my claim that strong epistemic beliefs are false is protected from worries which might arise if such beliefs were justified.

My Evolutionary Debunking Argument for strong epistemic beliefs is the following:

Causal premise. The existence of strong epistemic beliefs in humans is explained by the adaptiveness of such beliefs.

Epistemic premise. Strong epistemic beliefs are produced by an off-track process (by mechanisms functioning to produce useful beliefs).

Therefore,

Strong epistemic beliefs are unjustified.

Let us now look at the causal premise and the epistemic premise in turn.

6.4.1 Causal Premise

I am a cognitivist about strong epistemic beliefs, that is, I think that they express propositions. However, I claim that they are all false. An explanation can be given for why we have such beliefs by appeal to the biological usefulness of them. Strong epistemic beliefs serve a biological advantage which cannot be secured by Transparency to truth in doxastic deliberation, the Uncontrollability Thesis, and (if we had them) weak epistemic beliefs. If we did not think our beliefs had a standard of correctness irreducible to biology, and were governed by *sui generis* irreducible categorical epistemic norms, we would have less true beliefs than we do now, *even if* our doxastic deliberation were characterised by Transparency, and thus the Uncontrollability Thesis held. These features of belief, I suggest, are not enough to ensure that ‘most [of our] beliefs are true’ (Dennett 1987: 19, fn. 1), as they are.

It is in the following way that I think strong epistemic beliefs are biologically useful, and it is because they are useful in the following way, that the mechanisms which produce them have proliferated. Recall, such mechanisms function to produce useful beliefs, though not useful as an

approximation to truth, but rather useful with respect to self-directed effective functioning. Having true beliefs, that is, having beliefs which reflect Weak Epistemic Normativity, would result in less true beliefs than we would have were we to have false beliefs reflecting strong Epistemic Normativity. If we recognised that the only normativity attaching to our belief forming and ascribing practices was normativity rooted in biology, we would be less inclined to form beliefs in accordance with the standards provided by epistemic norms derivable from biology.

Compare the case of belief forming practices to the following two examples: biologically speaking, we are *supposed to* reproduce our kind. We recognise that biology lays down this expectation (in a very weak, non-intentional sense of ‘expectation’). Yet many of us choose not to reproduce. We do not allow the standards laid down by biology to motivate us inescapably to act in accordance with them. Similarly, we might recognise that one of the functions of the female breast is to feed infants, and yet some mothers choose not to breastfeed. There is a biological imperative to reproduce, and to breastfeed, but this is something we can turn our backs on. Perhaps if we took there to be *sui generis* irreducible categorical procreation or breastfeeding norms, generated from procreation or breastfeeding normativity, we would be less inclined to choose not to act in accordance with the standards laid down by biology for these activities.

The epistemic case looks different. Suppose that I am right, that biologically speaking, we are supposed to believe truly (in circumstances Normal for the performance of proper function one), and that there are categorical norms derivable from biology about how we might go about producing beliefs in line with the standard of correctness laid down by biology. The way we think about epistemic standards, norms, and obligations does not reflect the truth about them. My claim is that if we had true beliefs about the source of epistemic normativity (its being reducible to biological normativity), we may, as with the reproduction and breastfeeding case, be less inclined to form beliefs in line with epistemic norms. Having strong epistemic beliefs, that is, beliefs at the wrong strength with respect to the normativity involved in belief, makes it psychologically more difficult to turn our backs on the kinds of behaviours and belief-forming practices epistemic norms prescribe.

To see this, imagine a world in which everything is the same as it is in the actual world, except there are no strong epistemic beliefs. The belief formation of creatures in this world is,

though, characterised by Transparency and they thus cannot bring about beliefs at will.⁶² We might also suppose that they have weak epistemic beliefs (which are, on my account, true), regarding the standard of correctness laid down for belief by biology, and the norms reducible to biology relating to how best to meet that standard. Would these believers be less successful believers than us, that is, would they have fewer true beliefs, less justified beliefs, less rational beliefs (where these last two properties are understood descriptively)? I think that they would. It is true that they would not be able to have beliefs that *p merely* for reasons not related to the truth of *p*—this is secured by Transparency. It is true that they would also recognise that true beliefs were *correct* and false beliefs *incorrect*, and that there are norms reducible to biology which direct belief forming practices towards true beliefs. However, recognising that there are standards laid down by biology with respect to our belief forming practices, does not in any way secure conforming to those standards. We could recognise that there is a sense in which beliefs are supposed to be true, in the same way that there is a sense in which we should reproduce, or breastfeed, but we may not be any more motivated to mould our belief forming practices in line with what biology dictates than we are motivated to reproduce or breastfeed. What this means is that we may well be psychologically more able to turn our backs on epistemic norms than we are now. Perhaps then there would be beliefs that *p* based on very little evidence, even beliefs which the believer had more evidence against than for, and this could be a systematic and regular occurrence. This is consistent with doxastic deliberation being characterised by Transparency, the Uncontrollability Thesis being true, and our having weak epistemic beliefs. Evolution characterising our belief formation by Transparency, only guarantees that we cannot—deliberatively—be swayed in our belief formation by considerations which are not epistemic in kind. Our having weak epistemic beliefs which reflect the true nature of Epistemic Normativity only guarantees that we recognise in a weak sense, that belief has a standard of correctness and there are ways to go about meeting that standard. What is not guaranteed by the presence of these features is that one believes that *p* on *sufficient* or *reasonable* evidence, or that we feel motivated to respond to epistemic norms. It does not guarantee that ‘we *care* about truth and

62 One might think that human children are rather like these creatures. There may be empirical evidence which supports the view that children do not have epistemic beliefs. If this is right, then the evolutionary explanation of epistemic normativity gets pushed back. Now, the structuring cause of our strong epistemic beliefs might be thought to be education, it is education which endows us with an ‘epistemic sense’. However, even if children do not have strong epistemic beliefs, that does not mean that it is education which gives them such beliefs. They may naturally develop as part of their cognitive development—some stuff just does not show up until later. So this would not mean that strong epistemic beliefs are brought about by education. Thanks to Will McNeill, Paul Noordhof, and Tom Stoneham for discussion on this point.

rationality' or that 'we feel embarrassed when we catch ourselves wishfully thinking' (Price manuscript: 12). However, belief for us is not like it is for these creatures. We often scrutinise evidence, consider evidence from both sides, refrain from belief until we feel our evidence is *sufficient*, scold ourselves for not conforming to epistemic norms, and so on. I think that our falsely 'recognising' a *sui generis* irreducible normativity attaching to belief—Strong Epistemic Normativity—is in part responsible for this. It might be that confidence thresholds are kept high in some cases by strong epistemic beliefs, and we appeal to strong epistemic norms when we justify our beliefs in high stake cases.

So there is room for a biological explanation of the illusion of strong Epistemic Normativity, given that a biological account of Transparency, the Uncontrollability Thesis, and our having weak epistemic beliefs, is not (or would not be) sufficient to explain everything about the (contingent) nature of our beliefs. Natural selection is to be credited for more than just making our doxastic deliberation transparent to truth considerations. As Railton puts it:

[e]volution almost certainly selected our cognitive faculties to do more than represent accurately, and we certainly value, even value *epistemically*, more than truth: we also care about whether our beliefs are justified, coherent, manageable, applicable, informative, confident, explanatory, fruitful, etc. (Railton 1994: 75)

Our caring about this, in the way that we do, is something which would not be secured by having weak epistemic beliefs. It is secured by our mistaking biological normativity for *sui generis* epistemic normativity. We have seen why—from a biological point of view—the illusion of strong Epistemic Normativity, that is, having epistemic beliefs at a particular strength might be a good idea. So how does my account explain the presence of such beliefs?

In order to make our beliefs more likely to perform their derived proper function of being true, our biological history has ensured we have thoughts about correct, incorrect, rational, irrational, justified, unjustified (and so on) belief, where we do not take such evaluations to be reducible to biology. That we hold our beliefs up to *epistemic* standards (not merely standards laid down by biology); that we feel we ought to have *sufficient* or *good* evidence for our beliefs and that others *ought* to apply the same standards to their beliefs. We have evolved dispositions which incline us to judge our beliefs and the beliefs of others in these ways. Our feeling it appropriate to hold our beliefs to epistemic norms—to standards of production—is a way of making it more likely that our beliefs have true contents. If we accept this account, the illusion of Strong Epistemic Normativity is explained. We can account for the illusion of Strong Epistemic

Normativity by appealing to this biological account of strong epistemic beliefs, without having to posit irreducible normativity, and violate Hazlett's two naturalistic constraints.

A parallel can be drawn here with what the evolutionary metaethicist says about moral beliefs and moral properties. The claim is that evolution has insured we have thoughts (and beliefs) about right and wrong in order to make us cooperate with one another in line with our biological ends (Ruse 1986: 99). Ethical claims are 'pressed upon our thinking because of their adaptive value' (Ruse 1986: 102). So just as we can offer a causal argument citing the etiology of our ethical sentiments, so too can we offer a similar argument as to the etiology of our strong epistemic beliefs, both kinds of belief come out as 'illusion[s]', which have been 'fashioned and maintained by natural selection in order to promote individual reproduction' (Ruse 1986: 102). There might be moral standards laid down by biology, but if we took moral normativity to be reducible in this way, we might be less inclined to act morally. If one thought that the evolutionary metaethicist's account was right, this would be another example of our mechanisms for belief-production producing devices which are *useful*, instead of true.⁶³

For the evolutionary metaethicist, there still remains a degree of freedom as to our moral actions, that is, we 'are not blindly locked into our courses of action like robots. We are inclined to behave morally but not predestined to such a policy' (Ruse 1986: 99). In contrast, we cannot answer the question *whether p* without believing that *p* is true. On my account the reason we do not enjoy this kind of freedom when we form beliefs is grounded in causal facts which give rise to the disposition to be moved from the question *whether to believe that p* to the question *whether p is true*, the causal facts which realise Transparency. Such causal facts are secured by natural selection.

However, we have a kind of epistemic freedom in that we can hold ourselves to more and less stringent evidential requirements, sometimes depending on the costs of getting something wrong. So though we are inclined to look for evidence for our beliefs, to believe only upon what we take to be sufficient evidence, and so on, we have some freedom in this regard. This freedom comes from the different thresholds for evidence we may have, which are affected by practical considerations, as discussed in chapter two (§2.3.1.2).

The causal premise of my Evolutionary Debunking Argument appeals to biological history to claim that our strong epistemic beliefs are instrumentally valuable insofar as the beliefs they sanction (i.e., the ones which meet our standards of epistemic evaluation), are instrumentally

63 I judged the case of moral beliefs to be too controversial to include in my earlier list of beliefs produced on the basis of a systematic doxastic bias (though I think that they do represent such a case).

valuable. The casual account of these kinds of belief explains why we are disposed to have them.

6.4.2 Epistemic Premise

The causal premise alone does not show that our strong epistemic beliefs are unjustified. Even if strong epistemic beliefs are adaptive, that does not demonstrate that there is only Weak Epistemic Normativity, and that its stronger counterpart is illusory. In his discussion of evolutionary metaethics, Robert Nozick claims that showing certain beliefs to be adaptive ‘does not undercut the possible role of ethical facts in the origin of ethical behaviour’ (Nozick 1983: 344). We can give an evolutionary explanation of why a given trait has spread in a population, but within such an explanation certain corresponding truths enter. The behaviour is adaptive because the beliefs on which it is based are true (Nozick 1983: 345). So we cannot go from an explanation of strong epistemic beliefs as adaptive to the claim that those beliefs are unjustified. They might be adaptive because they are responsive to genuine irreducible *sui generis* epistemic normativity which attaches only to belief.

We need a further argument for the claim that the causal story for the existence of our strong epistemic beliefs is an off-track one, that is, it is one which does not track truth. My claim now is that conditions for the performance of proper function two are abnormal conditions for the performance of proper function one. The performance of function two does not track the truth. I have argued that Weak Epistemic Normativity is true, it follows from this that strong epistemic beliefs are false. My explanation for why we nevertheless have them is that they are products of our mechanisms for belief-production performing the function of producing useful beliefs. This means that the process by which we come to have these beliefs is a process which is off-track. This means that strong epistemic beliefs are unjustified.

Strong epistemic beliefs are produced in line with the performance of the proper function of producing useful beliefs. On my account, the kinds of beliefs these mechanisms produce are those which are *useful* but are also, usually, *false*. We saw earlier that examples include beliefs based on self-enhancement bias, partiality bias, and self-deception. Our mechanisms for belief-production, when performing their function of producing useful beliefs, Normally (and normally) produce useful *false* beliefs. At the very least, this gets us the claim that our strong epistemic beliefs are *unjustified*; they are produced by mechanisms which are not tracking truth, and the products of these mechanisms, in other cases, are usually *false*. If this is right, we have an additional reason for thinking that not only are strong epistemic beliefs are *unjustified*, but that they are also false. The additional reason which we get from my account of the production of strong epistemic beliefs, is that the mechanisms which produce them are just *not interested* in

truth, it is not something which plays a role in the production of beliefs. I think that this is a reason to think that strong epistemic beliefs are not merely *unjustified* beliefs, but also *false* ones. Considerations of theoretical parsimony may also support this suggestion: why suppose that the truth value of strong epistemic beliefs differs from *all other beliefs* produced in line with this function? To think this, one would have to have a reason for thinking that strong epistemic beliefs bucked the trend, and such a reason would also come at the expense of metaphysical queerness (§6.2.2). However, if we have *no reason* to think that they do—and I think that we do not—the claim that strong epistemic beliefs are *false* rather than merely *unjustified* starts to look highly plausible (and supported by my arguments for the claim that weak epistemic beliefs are true). As I said earlier though, the function of my Evolutionary Debunking Argument is only to demonstrate that strong epistemic beliefs are unjustified, but I do think the epistemic premise of my Evolutionary Debunking Argument does lean towards the stronger conclusion that they are false. If the claim that strong epistemic beliefs are false stands (from Weak Epistemic Normativity being true, plus considerations of queerness, and my Evolutionary Debunking Argument suggesting that they are unjustified), I can now offer an error theory regarding these beliefs.

6.5 An Error Theory for Strong Epistemic Beliefs

So far I have argued that our strong epistemic beliefs are false, which follows from Weak Epistemic Normativity being true, and is supported by consideration of queerness, and protected by my Evolutionary Debunking Argument which pre-empts the claim that such beliefs are justified. My argument that weak epistemic beliefs are true and my Evolutionary Debunking Argument for strong epistemic beliefs being unjustified supports an error theory about these beliefs, which reveals that epistemic properties and epistemic reasons do not have all of the features we take them to have. Olson, a proponent of an error theory of epistemic beliefs, claims that ‘no first-order epistemic claim is true’, where such claims are those which state that there are epistemic reasons to believe or not to believe a given proposition, or that believing a given proposition is permissible or impermissible, and so on (Olson 2011: 77). He locates the error in epistemic discourse in supposing that there are categorical reasons to believe—the error theorist can talk about evidence for beliefs, but never categorical reasons for beliefs (Olson 2011: 88). Olson objects to the existence of categorical norms, but allows for hypothetical ones. His claim is that if epistemic norms were categorical ones, they would be ‘intrinsically prescriptive’, which is a metaphysically queer feature (Olson 2011: 81). My claim is different: I allow for categorical norms which are reducible to biology: biology lays down a standard for

beliefs from which they can deviate. We make a mistake then in thinking that epistemic norms indicate a *sui generis* irreducible normativity special to belief.

Olson claims that epistemic norms are hypothetical ones ‘since these norms have reason-giving force for some agents only in virtue of their roles or the activities they are engaged in’ (Olson 2011: 84). Terms like ‘responsible’ in epistemic discourse are used by the error theorist descriptively to indicate that an agent meets the standard for being a responsible believer (where this standard can also be understood in descriptive terms) (Olson 2011: 89). The error theorist is able to keep hold of ‘standards of epistemic merit and demerit’ by recasting claims in these terms as purely descriptive (Olson 2011: 90).

Olson’s account is right, insofar as it denies *irreducible* categorical epistemic normativity. However, as it stands it is problematic because his conception of epistemic normativity as hypothetical gets the motivational and phenomenological profile of epistemic discourse completely wrong. He claims that epistemic reasons are non-categorical reasons. This does not align well with how we treat such reasons:

[w]e certainly *treat* epistemic reasons as though they are categorical reasons in the course of our ordinary practice. Moreover, we treat epistemic reasons in this way from both the first- and third-person perspectives. That is, one treats epistemic reasons as categorical reasons both in offering such reasons to others as well as in responding to such reasons in the course of one’s own theoretical deliberations. (Kelly 2003: 621)

Also, we might think that if epistemic reasons were non-categorical, we might be able to weigh them. For example, if p ’s being true gives me an epistemic reason to believe that p , where that reason is a non-categorical reason, I ought to be able to weigh that reason against other reasons I might have not to believe that p . But I cannot do this (§2.3). Further, we think that we can have epistemic reasons to believe propositions even if ‘one’s believing those propositions holds no promise of advancing any goal which one actually possesses’ (Kelly 2003: 630). We can have epistemic reasons to believe propositions which are ‘not contingent upon whether I care about believing what is true’ (Cuneo 2007: 59). Olson is aware of this problem and goes some way towards answering it. He suggests that:

explicit judgements to the effect that one’s evidence strongly supports p typically result in belief that p because, typically, when we assess evidence for and against p we do so because we recognize non-categorical reasons to have true beliefs as to whether p ; we want to know whether p , we are interested in whether p , and we believe that were we to believe in accordance with evidence with respect to whether p we would come to know whether p . (Olson 2011: 86)

This response will not do, as it narrows the scope of cases to which we think epistemic reasons have categorical force. In cases in which we are in the game of wanting to form a belief about p , and so we are therefore interested in evidence for and against p , Olson's story goes through. So when a subject has evidence which strongly supports p , she believes that p because when she is engaged in assessing evidence for p , she might take herself to have a non-categorical reason to form a true belief about p . Olson might have drawn an analogy here with a case of clear hypothetical normativity and reasons. If I am playing chess and recognize that moving my rook four spaces to the left will result in putting my opponent into checkmate, this typically results in my moving my rook four spaces to the left, because, typically, when we assess possible moves in a chess game, we do so because we recognize hypothetical reasons to win the chess game, and we believe that were we to act in accordance with our findings, we would win the chess game.

However, these kinds of cases are not exhaustive. The kinds of cases Kelly and Cuneo have in mind are not those in which we 'recognize non-categorical reasons to have true beliefs as to whether p ' (Olson 2011: 86). Rather they are cases in which I have no interest in whether p is true, no goal of mine is advanced by coming to a belief on the matter of p , *and yet*, if I have evidence that p is true, I *nevertheless* take myself to have an epistemic reason to believe that p . This description is coherent, unlike the analogue in the chess case. If I have no interest in winning a game of chess, if I have evidence that moving my rook four places to the left will put my opponent into check mate, it is not the case that I *nevertheless* take myself to have a reason to move my rook four places to the left. The categorical force of epistemic reasons then, cannot be accommodated by an appeal to cases in which we recognize non-categorical reasons to form a belief about p , because this force is present in the absence of such reasons:

[o]n the usual view of things, two agents in the same epistemic situation (same evidence, same background beliefs) would have the same reasons for believing any given propositions, regardless of possible differences in their personal goals. (Railton 1997: 53)

We can also look at cases in which someone has non-epistemic reasons *not* to form the belief that p , but we still take them to have an *epistemic* reason to form that belief. Papineau imagines a subject who 'realises that he is likely to be upset if he learns about the real probability of his developing cancer, and so arranges to avoid any evidence that might undermine his sanguine belief that this probability is low' (Papineau 1999: 24). Papineau asks whether this subject, and similar subjects, are acting wrongly. His answer is that given that, *ex hypothesi*, such subjects do not want their beliefs to be true, it is 'not obvious' that there is another sense in which we might

think they are acting wrongly (Papineau 1999: 24). However, this again is not faithful to the way we *think* about epistemic normativity. There *is* a sense in which the subject is making a mistake, and acting wrongly, he is not complying with epistemic norms, which we take to be categorical and *sui generis*, ‘there is *a sense* in which these people are proceeding improperly, namely, the *epistemic* sense’ (Hazlett 2013: 173).

Olson’s account needs to be supplemented in order to deal with this problem. He has given a purely negative account by claiming that there are no categorical epistemic reasons or norms, but offered no indication of why he pushes an error theory, nor an explanation for the illusion of strong Epistemic Normativity.⁶⁴ From Olson then, we do not learn why we have false strong epistemic beliefs. My account can supplement Olson’s in these ways—it can explain why we *mistakenly* take epistemic reasons and norms to have a stronger status than they do.

The reason we have false strong epistemic beliefs is because it is adaptive to act or believe in line with what they license, and this explains Kelly’s and Cuneo’s observations.

6.6 Conclusion

In conclusion, I considered the objection that truth is not the only goal when it comes to the fixation of belief, and that this is indicated by the presence of systematic doxastic bias. I argued that we need to recognise a second function proper to our mechanisms for belief-production: the function of producing useful beliefs, understood as self-directed effective functioning beliefs.

I distinguished between Strong and Weak Epistemic Normativity and gave an explanation of the latter by appeal to the function of producing true beliefs. I discussed worries about queerness with respect to epistemic properties to support my claim that strong epistemic beliefs are false. I suggested that Strong Epistemic Normativity is not a feature which connects belief with truth, but is nonetheless a feature of belief which requires an explanation. I sought to provide an explanation by appeal to the function of producing useful beliefs. I then gave an Evolutionary Debunking Argument which showed that strong epistemic beliefs are unjustified. Finally, I argued that there are categorical epistemic norms, but they are reducible to biological norms. I suggested that Olson’s account gets the motivational profile of epistemic discourse wrong, and referred to Kelly’s and Cuneo’s claims that we treat epistemic reasons as categorical.

⁶⁴ That is not to suggest that he should have. Rather my claim is that the objection can be answered by supplementing the error theory with my account of the casual story of the illusion of Strong Epistemic Normativity.

I argued that Olson's response to this problem failed, and that my account is able to accommodate the way we treat epistemic norms. I conclude then, that the presence of Weak Epistemic Normativity, as well as the illusion of Strong Epistemic Normativity can be explained by my account of belief.

CONCLUSION

In conclusion, the three features of belief which indicate a connection between belief and truth place an explanatory burden on the belief theorist. My account meets this burden.

I began by defending the claim that what is essential to belief is its motivational role, that is, it is necessary and sufficient for something's being a belief that it has the appropriate motivational role. It is this which is able to distinguish beliefs from other attitudes, and I took the motivational role forward in the thesis as offering a necessary and sufficient condition for some attitude being a belief. I defended this claim from the objection that other attitudes play the motivational role specified for belief, and so something else is required to demarcate belief from these attitudes (it is in part this worry which motivated the teleological and normative accounts discussed in chapters two and three). I responded to this by claiming that though other attitudes can play the role of belief, belief is distinguished from those other attitudes by its being able to play the role *by itself* and *across contexts*. However, I argued that though the motivational account can answer this objection, the account cannot explain belief's link to truth, and so something else is required.

In chapter two I discussed the teleological account of belief. I argued that the account fails to capture our relatively passive commitment to truth suggestive of biological function, instead casting it as something towards which the subject is aimed. I argued that the account is not able to explain the link between belief and truth, in virtue of aims not being at the required strength to do the explanatory work. Further, I argued that two objections raised to the account remain unanswered. It is for these reasons that I suggested that the account ought to be rejected.

In chapter three I discussed the normative account of belief. I argued that the account must be couched in terms of permissibility and not obligation, because an account cast in terms of obligation required too much of believers. I then suggested that the account could not explain all of the features which indicate a link between belief and truth because the norm—now cast in terms of permissibility—is not at the required strength to do this explanatory work. I then considered the No Guidance Objection, the response to which demonstrated that the normative account makes a mistake in grounding the link to truth at an intentional level, rather than at the sub-intentional level secured by biology. For these reasons, I concluded that the normative account ought to be rejected.

From here I went on to develop my preferred account. In chapter four, I offered a development of the etiological account which sought to make clearer the causal-historical

relationship cited to hold between present day tokens and ancestral tokens of a given trait type. I argued that adopting this proposal would allow the etiologist to ground a notion of biological normativity, as well as explain malfunction. I also argued that my proposal could meet the objection that the etiological account cannot individuate trait types without circularity. I concluded that the etiological account is defensible and can ground a notion of biological normativity. I took this account of biological function forward into chapters five and six in which I developed my account of belief.

In chapter five I put forward the first part of my account. I claimed that our mechanisms for belief-production have the relational proper function of producing true beliefs, and that it is by appeal to this that we can explain Transparency and the Uncontrollability Thesis. My account could offer explanations of these phenomena which cited both triggering and structuring causes (whereas, we saw in chapters two and three that the teleological and normative accounts could only cite triggering causes). A consequence of my explanations was that these features were contingent ones grounded in biology. They were understood as features of the beliefs of *some believers*, rather than necessary features of *all beliefs*.

In chapter six I argued that we need to introduce a second function proper to our mechanisms for belief-production, since there are many cases of belief which suggest that the mechanisms which produce them are not geared towards truth. I suggested that this second function proper to our mechanisms for belief-production is the function of producing useful beliefs, though not useful as an approximation to truth, but rather, useful in assisting effective functioning of the believer. I argued that Weak Epistemic Normativity was explainable by appeal to the function of producing true beliefs and (the illusion of) Strong Epistemic Normativity was explainable by appeal to the function of producing useful beliefs.

In conclusion, the link between belief and truth as indicated by Transparency, the Uncontrollability Thesis and Weak Epistemic Normativity, is a contingent one explainable by appeal to the proper functions of our mechanisms for belief-production. We should not take beliefs to hold this link to truth as a matter of metaphysical necessity, this would be to project our particular circumstance as believers with a certain biological history, onto a claim about all beliefs. I conclude that the link between belief and truth can be explained by the biological functions of the mechanisms which produce beliefs in the actual world.

REFERENCES

- Alicke, Mark D. 1985: 'Global Self-Evaluation as Determined by the Desirability and Controllability of Trait Adjectives'. *Journal of Personality and Social Psychology*. Vol. 49, no. 6, pp. 1621–30.
- Allen, Colin and Bekoff, Mark 1995: 'Biological Function, Adaptation, and Natural Design'. *Philosophy of Science*. Vol. 62, no. 4, pp. 609–22.
- Armstrong, David M. 1973: *Belief, Truth and Knowledge*. London: Cambridge University Press.
- Bennett, Jonathan 1990: 'Why Is Belief Involuntary?' *Analysis*. Vol. 50, no. 2, pp. 87–107.
- Bigelow, John and Pargetter, Robert 1987: 'Functions'. *The Journal of Philosophy*. Vol. 84, no. 4, pp. 181–96.
- Boghossian, Paul 2003: 'The Normativity of Content'. *Philosophical Issues*. Vol. 13, no. 1, pp. 31–45.
- Bratman, Michael E. 1992: 'Practical Reasoning and Acceptance in a Context'. *Mind*. Vol. 101, no. 401, pp. 1–15.
- Brown, Johnathon D. 1986: 'Evaluations of Self and Others: Self-Enhancement Biases in Social Judgments'. *Social Cognition*. Vol. 4, no. 4, pp. 353–76.
- Buleandra, Andrei 2009: 'Doxastic Transparency and Prescriptivity'. *Dialectica*. Vol. 63, no. 3, pp. 325–32.
- Buller, David 1998: 'Etiological Theories of Function: A Geographical Survey'. *Biology and Philosophy*. Vol. 13, no. 4 pp. 505–27.
- Butler, Peter V. 2000: 'Reverse Othello Syndrome Subsequent to Traumatic Brain Injury'. *Psychiatry: Interpersonal and Biological Processes*. Vol. 63, no. 1, pp. 85–92.
- Bykvist, Krister and Hattiangadi, Anandi 2007: 'Does Thought Imply Ought?' *Analysis*. Vol. 67, no. 4, pp. 277–85.
- Chan, Timothy 2013: 'Introduction: Aiming at Truth'. In Chan, T (ed.) *The Aim of Belief*. USA: Oxford University Press, pp. 1–16.
- Cummins, Robert 1975: 'Functional Analysis'. *The Journal of Philosophy*. Vol. 72, no. 20, pp. 741–65.
- Cuneo, Terence 2007: *The Normative Web*. Oxford: Oxford University Press.
- Davidson, Donald 1987: 'Knowing One's Own Mind'. *Proceedings and Addresses of the American Philosophical Association*. Vol. 60, no. 3, pp. 441–58.
- 2001: 'Thought and Talk'. In *Inquires into Truth and Interpretation*. USA: Oxford University press, pp. 155–70.

- Davies, Paul Sheldon 1994: 'Trouble for Direct Proper Functions'. *Noûs*. Vol. 28, no. 3, pp. 363–81.
- 2000: 'Malfunctions'. *Biology and Philosophy*. Vol. 15, no. 1. pp. 19–38.
- 2001: *Norms of Nature: Naturalism and the Nature of Functions*. Cambridge Massachusetts: MIT Press.
- Dawkins, Richard 2009: *The Greatest Show on Earth: The Evidence for Evolution*. United Kingdom: Bantam Press.
- Dennett, Daniel 1987: *The Intentional Stance*. USA: MIT Press.
- Dokic, Jerome and Engel, Pascal 2002: *Frank Ramsey: Truth and Success*. London: Routledge.
- Dretske, Fred 1988: *Explaining Behavior*. MIT: USA.
- 2004: 'Psychological vs. Biological Explanations of Behavior'. *Behavior and Philosophy*. Vol. 32, no. 1, pp. 167–77.
- Dunning, David, Meyerowitz, Judith A. and Holzberg, Amy D. 1989: 'Ambiguity and Self-Evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessments of Ability'. *Journal of Personality and Social Psychology*. Vol. 57, no. 6, pp. 1082–90.
- Engel, Pascal 2006: 'Epistemic Norms and Rationality'. In Strawinski, W., Grygianca, M., and Brodek, A. (eds.) *Mysli o Jezyku, nauce i wartosciach, Ksiega ofiarowana Jackowi Juliużowi Jadakiemu*, Warszawa: Semper: Springer. pp. 359–74.
- 2007: 'Belief and Normativity'. *Disputatio*. Vol. 2, no. 23, pp. 179–204.
- 2013: 'Doxastic Correctness'. *Proceedings of the Aristotelian Society Supplementary Volume*. Vol. 87, no. 1, pp. 199–216.
- Fales, Evan 2002: 'Darwin's Doubt, Calvin's Calvary'. In Bielby, James (ed.) *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*. USA: Cornell University Press, pp. 43–58.
- Feldman, Richard 1988a: 'Epistemic Obligations'. *Philosophical Perspectives*. Vol. 2, pp. 235–56.
- 1988b: 'Rationality, Reliability and Natural Selection'. *Philosophy of Science*. Vol. 55, no. 2, pp. 218–27.
- Fodor, Jerry 2002: 'Is Science Biologically Possible?' In Bielby, James (ed.) *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*. USA: Cornell University Press, pp. 30–42
- Frankish, Keith 2007: 'Deciding to Believe Again'. *Mind*. Vol. 116, no. 463, pp. 523–48.
- Gibbard, Allan 2005: 'Truth and Correct Belief'. *Philosophical Issues*. Vol. 15, no. 1, pp. 338–50.

- Gluer, Kathrin and Wikforss, Asa 2009: 'Against Content Normativity'. *Mind*. Vol. 118, No. 469, pp. 31–68.
- Godfrey-Smith, Peter 1994: 'A Modern History Theory of Functions'. *Noûs*. Vol. 28, no. 3, pp. 344–62.
- Green, Melanie C. and Brock, Timothy C. 2000: 'The Role of Transportation in the Persuasiveness of Public Narratives'. *Journal of Personality and Social Psychology*. Vol. 79, no. 5, pp. 701–21.
- Griffiths, Paul 1993: 'Functional Analysis and Proper Functions'. *The British Journal for the Philosophy of Science*. Vol. 44, no. 3, pp. 409–22.
- Grimm, Stephen 2009: 'Epistemic Normativity'. In Haddock, Adrian, Millar, Alan and Pritchard, Duncan (eds.) *Epistemic Value*. Oxford: Oxford University Press, pp. 243–64.
- Hardcastle, Valerie Gray 2002: 'On the Normativity of Functions'. In Arieu, Andre, Cummins, Robert and Perlman, Mark (eds.) *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford: Oxford University Press, pp. 144–56.
- Hazlett, Allan 2013: *A Luxury of the Understanding: On the Value of True Belief*. Oxford: Oxford University Press.
- Helgeson, Vicki S. and Taylor, Shelley E. 1993: 'Social Comparisons and Adjustment Among Cardiac Patients'. *Journal of Applied Social Psychology*. Vol. 23, no. 15, pp. 1171–95.
- Kelly, Thomas 2003: 'Epistemic Rationality as Instrumental Rationality: A Critique'. *Philosophy and Phenomenological Research*. Vol. 66, no. 3, pp. 612–40.
- Kahane, Guy 2011: 'Evolutionary Debunking Arguments'. *Nous*. Vol. 45, no. 1, pp. 103–25
- Kettlewell, Henry Bernard Davies: 1973: *The Evolution of Melanism: The Study of a Recurring Necessity*. Oxford: Oxford University Press.
- Kunda, Ziva 1990: 'The Case for Motivated Reasoning'. *Psychological Bulletin*. Vol. 108, no. 3, pp. 480–98.
- Lewinsohn, Peter M., Mischel, Walter, Chaplin, William, and Barton, Russell 1980: 'Self Competence and Depression: The Role of Illusory Self-Perceptions'. *Journal of Abnormal Psychology*. Vol. 89, no. 2, pp. 203–12.
- Loewer, Barry 1993: 'The Value of Truth'. *Philosophical Issues*. Vol. 4, pp. 265–80.
- Lynch, Michael 2004: *True to Life: Why Truth Matters*. USA: MIT Press.
- Macdonald, Graham and Papineau, David 2006: 'Introduction: Prospects and Problems for Teleosemantics'. In Macdonald, Graham and Papineau, David (eds.), *Teleosemantics*, Oxford: Oxford University Press, pp. 1–22.
- Mayo, Bernard 1964: 'Belief and Constraint'. *Proceedings of the Aristotelian Society*. Vol. 64, pp. 139–56.

- McHugh, Conor 2011: 'What Do We Aim At When We Believe?' *Dialectica*. Vol 65, no. 3, pp. 369–92.
- 2012: 'Beliefs and Aims'. *Philosophical Studies*. Vol. 160, no. 425, pp. 425–39.
- *forthcoming*: 'The Illusion of Exclusivity'. *European Journal of Philosophy*. DOI: 10.1111/ejop.12032
- McLaughlin, Peter 2001: *What Functions Explain*. United Kingdom: Cambridge University Press.
- Mele, Alfred 2000: *Self-Deception Unmasked*. USA: Princeton University Press.
- Millikan, Ruth Garrett 1984: *Language, Thought and Other Biological Categories*. USA: MIT Press.
- 1989a: 'An Ambiguity in the Notion "Function"'. *Biology and Philosophy*. Vol. 4, no. 2, pp. 172–76.
- 1989b: 'Biosemantics'. *The Journal of Philosophy*, Vol. 86, no. 6, pp. 281–97.
- 1989c: 'In Defense of Proper Functions'. *Philosophy of Science*, Vol. 56, no. 2, pp. 288–302.
- 1995a: 'Explanation in Biopsychology'. In *White Queen Psychology*. MIT: USA. pp. 171–92.
- 1995b: 'Naturalist Reflections on Knowledge'. In *White Queen Psychology*. MIT: USA. pp. 241–64.
- 1995c: 'Propensities, Exaptations, and the Brain'. In *White Queen Psychology*. MIT: USA. pp. 31–50.
- 1995d: 'Thoughts Without Laws'. In *White Queen Psychology*. MIT: USA. pp. 51–83.
- 1999: 'Historical Kinds and the "Special Sciences"'. *Philosophical Studies*, Vol. 95, no. 1/2, pp. 45–65.
- 2002: 'Biofunctions: Two Paradigms'. In Ariew, Andre, Cummins, Robert and Perlman, Mark (eds.) *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford: Oxford University Press, pp. 113–43.
- Nanay, Bence 2010: 'A Modal Theory of Function'. *The Journal of Philosophy*. Vol. 107, no. 8, pp. 412–31.
- 2012: 'Function Attributions Depend on the Explanatory Context: A Reply to Neander and Rosenberg's Reply to Nanay'. Vol. 109, no. 10, pp. 623–27.
- Neander, Karen 1991a: 'Functions as Selected Effects: The Conceptual Analyst's Defense'. *Philosophy of Science*, Vol. 58, no. 2, pp. 168–84.

----- 1991b: 'The Teleological Notion of "Function"'. *Australasian Journal of Philosophy*. Vol. 69, no. 4, pp. 454–68.

----- 2012: 'Teleological Theories of Mental Content'. *The Stanford Encyclopedia of Philosophy (Spring 2012 Edition)*, Edward N. Zalta (ed.). URL=
<http://plato.stanford.edu/archives/spr2012/entries/content-teleological>

Neander, Karen and Rosenburg, Alex 2012: 'Solving the Circularity Problem for Functions: A Response to Nanay'. *The Journal of Philosophy*. Vol. 109, no. 10, pp. 613–22.

Nichols, Shaun and Stich, Stephen 2003: *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Oxford University Press.

Noordhof, Paul 2001: 'Believe What You Want'. *Proceedings of the Aristotelian Society*, Vol. 101, no. 1, pp. 247–65.

----- 2003: 'Self-Deception, Interpretation and Consciousness'. *Philosophy and Phenomenological Research*. Vol. 67, no. 1, pp. 75–100.

Nozick, Robert 1983: *Philosophical Explanations*. USA: Belknap Press of Harvard University Press.

O'Brien, Lucy 2005: 'Imagination and the Motivational View of Belief'. *Analysis*. Vol. 65, no. 1, pp. 55–62.

Olson, Jonas 2011: 'Error Theory and Reasons for Belief'. In Reisner, Andrew and Steglich-Petersen, Asbjørn (eds.) *Reasons for Belief*. United Kingdom: Cambridge University Press, pp. 75–93.

Owens, David J 2003: 'Does Belief Have an Aim?' *Philosophical Studies*, Vol. 115, no. 3, pp. 283–305.

Papineau, David 1987: *Reality and Representation*. Great Britain Basil Blackwell Ltd.

----- 1991: 'Truth and Teleology'. In Knowles, Dudley (ed.) *Explanation and Its Limits*. USA: Cambridge University Press, pp. 21–44.

----- 1993: *Philosophical Naturalism*. Blackwell Publishers: Cambridge, Massachusetts.

----- 1999: 'Normativity and Judgement'. *Proceedings of the Aristotelian Society*, Vol. 73, no. 1, pp. 17–43.

----- 2001: 'The Status of Teleosemantics, or How to Stop Worrying about Swampman'. *Australasian Journal of Philosophy*. Vol. 79, no. 2, pp. 279–89.

----- 2013: 'There are No Norms of Belief'. In Chan, T (ed.) *The Aim of Belief*. USA: Oxford University Press, pp. 64–79.

Plantinga, Alvin 2002: 'Introduction: The Evolutionary Argument Against Naturalism: An Initial Statement of the Argument'. In Bielby, James (ed.) *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*. USA: Cornell University Press, pp. 1–12.

Price, Carolyn 1995: 'Functional Explanations and Natural Norms'. *Ratio*. Vol. 7, no. 2, pp. 143–60.

----- 2001: *Functions in Mind*. USA: Oxford University Press.

----- *Manuscript*: 'Belief and Truth: Aim or Function'.

Pritchard, Duncan 2011: 'What is the Swamping Problem?' In Reisner, Andrew, and Steglich-Petersen, Asbjørn (eds.) *Reasons for Belief*. USA: Cambridge University Press. pp. 244–59.

Quine, W. V. O 1985: 'Natural Kinds'. In Kornblith, Hilary (ed.) *Naturalizing Epistemology*. USA: MIT Press.

Ramsey, Frank P. 1931: *The Foundations of Mathematics and Other Logical Essays*. London: Routledge.

Ramsey, William 2002: 'Naturalism Defended'. In Bielby, James (ed.) *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*. USA: Cornell University Press, pp. 15–19.

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

Ruse, Michael 1986: 'Evolutionary Ethics: A Phoenix Arisen'. *Zygon*. Vol. 21, no. 1, pp. 95–112.

Sage, James 2004: 'Truth-Reliability and the Evolution of Human Cognitive Faculties'. *Philosophical Studies*. Vol. 177, no. 1/2, pp. 95–106.

Scott-Kakures, Dion 1993: 'On Belief and the Captivity of the Will'. *Philosophy and Phenomenological Research*. Vol. 53, no. 4, pp. 77–103.

Sedikides, Constantine 1993: 'Assessment, Enhancement, and Verification Determinants of the Self-Evaluation Process'. *Journal of Personality and Social Psychology*. Vol. 65, no. 2, pp. 317–38.

Shah, Nishi 2003: 'How Truth Governs Belief'. *The Philosophical Review*. Vol. 112, no. 4, pp. 447–82.

Shah, Nishi and Velleman, J. David 2005: 'Doxastic Deliberation'. *The Philosophical Review*, Vol. 11, no. 4, pp. 497–534.

Sober, Elliott 1984: *The Nature of Natural Selection*. USA: MIT Press.

Sorensen, R. 1988: *Blindspots*. New York: Clarendon Press.

- Sosa, Ernest 2011: *Knowing Full Well*. Princeton: Princeton University Press.
- Stalnaker, Robert C. 1987: *Inquiry*. USA: MIT Press.
- Stanley, Jason 2005: *Knowledge and Practical Interests*. USA: Oxford University Press.
- Steglich–Petersen, Asbjørn 2006: ‘No Norm Needed: On the Aim of Belief’. *The Philosophical Quarterly*, Vol. 56, no. 225, pp. 499–516.
- 2008: ‘Does Doxastic Transparency Support Evidentialism?’ *Dialectica*, Vol. 62, no. 4, pp. 541–47.
- 2009: ‘Weighing the Aim of Belief’. *Philosophical Studies*, Vol. 145, no. 3, pp. 395–405.
- Stephens, Christopher L. 2001: ‘When Is It Selectively Advantageous to Have True Beliefs?’ *Philosophical Studies*. Vol. 105, no. 2, pp. 161–89.
- Stich, Stephen 1990: *The Fragmentation of Reason*. MIT: USA.
- Stuart–Fox, Devi and Moussalli, Adnan 2008: ‘Selection for Social Signalling Drives the Evolution of Chameleon Colour Change’. *PLoS Biology*. Vol. 6, no. 1, pp. 22–9.
- Sullivan–Bissett, Ema and Noordhof, Paul 2013: ‘A Defence of Owens’ Exclusivity Objection to Beliefs Having Aims’. *Philosophical Studies*. Vol. 167, no. 2, pp. 453–7.
- Taylor, Shelley E., Lerner, Jennifer S., Sherman, David K., Sage, Rebecca M., and McDowell, Nina K. 2003: ‘Are Self–Enhancing Cognitions Associated with Healthy or Unhealthy Biological Profiles?’ *Journal of Personality and Social Psychology*. Vol. 85, no. 4, pp. 605–15.
- Trivers, Robert 2000: ‘The Elements of a Scientific Theory of Self–Deception.’ *Annals NY Acad Sciences*. Vol. 907, pp. 114–31
- 2011: *Deceit and Self–Deception: Fooling Yourself the Better to Fool Others*. Great Britain: The Penguin Group.
- Trope, Yaacov and Liberman, Akiva 1996: ‘Social Hypothesis Testing: Cognitive and Motivational Mechanisms’. In Higgins, Edward and Kruglanski, Arie (eds.) *Social Psychology: Handbook of Basic Principles*. New York: Guildford Press, pp. 239–70.
- van Leeuwen, Neil 2007a: ‘The Product of Self–Deception’. *Erkenntnis*. Vol. 67, pp. 419–37.
- 2007b: ‘The Spandrels of Self–Deception.’ *Philosophical Psychology*. Vol. 20, no. 3, pp. 329–48
- 2008: ‘Finite Rational Self–Deceivers’. *Philosophical Studies*. Vol. 139, no. 2, pp.191–208
- 2009: ‘The Motivational Role of Belief’. *Philosophical Papers*. Vol. 38, no. 2, pp. 219–46.

- Velleman, David J. 2000: *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- Wedgwood, Ralph 2002: 'The Aim of Belief'. *Philosophical Perspectives*, Vol. 16, no. 16, pp. 267–97.
- Whiting, Daniel 2010: 'Should I Believe the Truth?' *Dialectica*. Vol. 64, no. 2, pp. 213–24.
- Whyte, J.T. 1990: 'Success Semantics'. *Analysis*. Vol. 50, no. 3, pp.149–57.
- Williams, Bernard 1970: 'Deciding to Believe', in his *Problems of the Self*. Cambridge: Cambridge University press, pp. 136–51.
- Wimsatt, William C. 1972: 'Teleology and the Logical Structure of Function Statements'. *Studies in the History of Philosophy of Science*. Vol. 3, no. 1, pp. 1–80.
- Winters, Barbara 1979: 'Believing at Will'. *The Journal of Philosophy*, Vol. 76, no. 5, pp. 243–56
- Wouters, Arno 2005: 'The Function Debate in Philosophy'. *Acta Biotheoretica.*, Vol. 53, no. 2, pp. 123–51.
- Wright, Larry 1973: 'Functions'. *Philosophical Review*. Vol. 82, no. 2, pp. 139–68.
- Wright, Larry 1976: *Teleological Explanations: An Etiological Analysis of Goals and Functions*. London: University of California Press.