

ON BELIEF:
AIMS, NORMS, AND FUNCTIONS

ATKINSON CHRISTOPHER JOHN

PHD

LINGNAN UNIVERSITY

2017

ON BELIEF:
AIMS, NORMS, AND FUNCTIONS

by
ATKINSON Christopher John

A thesis
submitted in partial fulfillment
of the requirements for the
Degree of Doctor of Philosophy in Philosophy

Lingnan University

2017

ABSTRACT

On Belief: Aims, Norms, and Functions

by

ATKINSON Christopher John

Doctor of Philosophy

In this thesis, I explore answers to three central questions: (i) what are beliefs, (ii) why do we have them, and (iii) how should we interpret doxastic correctness, the principle that it is correct to believe that p if and only if p ? The first question has a long history in philosophy of mind, and in various forms can be dated at least as far back as David Hume. For that reason, I refer to the problem as *Hume's Problem*. As I interpret the question, the main difficulty with accounting for what beliefs are, is in distinguishing them from forms of *acceptance*, where acceptances are understood as *regarding-as-true* attitudes—these include assuming, supposing, guessing, and (propositional) imagining. I take this technical use of ‘acceptance’ from Velleman (2000). The second question I interpret as a biological question about why we, as organisms with beliefs, have beliefs. Answering this question depends on how we answer the first question, in the sense that, in explaining why we have beliefs, we are explaining why we have an attitude that meets the conditions that we set for beliefs to meet. And the third question is largely about the important relation between belief and truth; why is it that, as believers, we emphasise the apparent platitude that a belief is correct if and only if it is true?

To address these questions, I consider three different theoretical approaches to understanding beliefs. Specifically, I discuss teleological theories, normative theories, and a functional theory. For various reasons, I argue that teleological accounts fail to provide satisfactory answers to the three central questions (Part I); that normative accounts also fail to answer our central questions (Part II); but that a functional account, appropriately understood, can provide answers to these questions (Part III). In particular, I argue for what I call the *doxastic effects thesis*, which defines belief according to the effects beliefs have (or their outputs); and I propose that we interpret the components of this thesis as *functional statements*. The doxastic effects thesis allows us to answer Hume's Problem, by proposing necessary and sufficient conditions for beliefs to meet; and interpreting these conditions as functions allows us to explain, at least in part, why we have beliefs. Concerning doxastic correctness, I argue that our commitment to the principle has arisen as a social construct, and therefore should be given (what I call) a *thin* reading; as opposed to the *substantive* reading that we get on the teleological and normative accounts, such that doxastic correctness states an essential fact about *belief*.

Finally, in Part IV, I extend the first question, what beliefs are, to a third doxastic attitude. Namely, suspended belief (suspension). I argue that one further advantage of my functional theory of belief is that it can account for suspension *as* a doxastic attitude, unlike the teleological and normative alternatives.

DECLARATION

I declare that this is an original work based primarily on my own research, and I warrant that all citations of previous research, published or unpublished, have been duly acknowledged.

ATKINSON Christopher John

CONTENTS

LIST OF FIGURES	iii
ACKNOWLEDGEMENTS	iv
DEDICATION	v
INTRODUCTION	iv
THESIS STRUCTURE	5
PART I: AIMS	12
1. Teleological Theories of Belief	12
1.1. The Sufficiency of the Truth-Aim.....	17
1.2. The Necessity of the Truth-Aim	25
2. The Exclusivity Objection.....	36
2.1. Explaining the Exclusivity Objection	37
2.1.1. Can the Aim of Belief be Weighed?	39
2.1.2. Rejecting Exclusivity: McHugh’s Pragmatism.....	43
2.1.2.1. McHugh’s Teleological Thesis	44
2.1.2.2. McHugh’s Discretion Principle.....	47
2.1.2.3. The Conflict.....	48
3. Summary of ‘Aims’.....	52
PART II: NORMS	55
4. Normative Theories of Belief	55
5. Objections from Normative Impotence	59
5.1. The No-Guidance Objection.....	59
5.1.1. Guidance by Ideal	66
5.2. The Regress of Rules Problem	74
5.2.1. Engel’s Reply	77
6. Summary of ‘Norms’	79
PART III: FUNCTIONS	82
7. The Doxastic Effects Theory of Belief	82
7.1. The Motivational Role Thesis.....	85
7.1.1. The Motivational Role of Imagining (and other Acceptances)	90
7.2. The Fundamentality Thesis.....	93
7.3. Solving Hume’s Problem	97

8. Etiological vs. Systemic Functions	98
8.1. Etiological and Systemic Functions	100
9. The Project of Functional Theorists	107
9.1. Conceptual Analysis.....	108
9.2. Theoretical Definition	115
10. The Normativity of Functions	120
10.1. Intuiting Natural Norms	123
10.2. Etiological Functions and Normativity	126
11. The Explanatory Power of Functions	134
12. A (Systemic) Functional Theory of Belief.....	140
12.1. A Systemic Analysis of the Motivational Role Thesis	140
12.2. A Systemic Analysis of the Fundamentality Thesis	145
12.3. Why We Have Beliefs.....	148
13. Doxastic Correctness	151
13.1. True Belief as Correct Belief.....	152
13.2. Non-Defective False Beliefs	154
14. Summary of ‘Functions’	157
PART IV: SUSPENDED BELIEF	161
15. Theories of Suspended Belief	161
15.1. Against a Teleological Reading of Suspension.....	162
15.2. Against a Normative Reading of Suspension.....	165
15.3. A Doxastic Effects Theory of Suspension	168
15.3.1. The Motivational Properties of Suspension	168
15.3.2. The Fundamentality Properties of Suspension	169
16. Summary of ‘Suspended Belief’	170
CONCLUSION	172
BIBLIOGRAPHY	175

List of Figures

Figure 1	Systemic Functions and the Heart	137
Figure 2	Systemic Functions and the Motivational Thesis	141
Figure 3	Systemic Functions and True Belief	143
Figure 4	Systemic Functions and the Fundamentality Thesis	147

Acknowledgements

I'd like to thank my supervisors, Darrell Rowbottom and Jennifer Nado, for their excellent guidance throughout this process. I'd also like to thank everyone else in the Philosophy Department at Lingnan, for helpful philosophical discussions over the years, and, at times, interesting nights out.

For Mum

Introduction

Almost all of us admit to having beliefs. We believe in gods, in love, in peace, in justice, in hate, in revenge; we believe in our abilities, our career prospects, and in the potentials of our children; and we typically have many more prosaic beliefs: that we will have dinner tonight, go swimming tomorrow, and probably won't win the lottery this weekend. Yet trying to give a precise account of what beliefs *are* is a task fraught with difficulties. Perhaps the most common assumption about beliefs is that they are *thoughts* that we *take to be true*. This, at least, is what I found in a quick (unscientific) survey of non-philosophers. So, when I say 'I believe that I am a philosopher', I mean something like 'I think it is true that I am a philosopher'. However, while this account touches on an important relation—that between belief and truth—it does not really take us very far.

From one perspective, it is patently uninformative. If we take the statement

I think that p is true

to be a simple rephrasing of

I believe that p is true

then the former holds no new information about belief. Believing that p just is the same as thinking that p is true. But what, we might ask, is thinking that p is true? If we answer 'believing that p ' then of course we end up in a circle that still leaves us wondering what beliefs are. However, this perspective is not quite right; we need to look at things differently. Some information *is* contained in the observation that believing is the same as thinking true, although it still only takes us half a step of the way.

To see what I mean, we need only recognise that the predicate 'is true' is redundant in the above *belief statement*, but not in the *thought statement*. To clarify, consider whether the following two statements are equivalent:

I think that p

and

I believe that p .

Clearly they are not—we can think that p without believing that p . In this sense, *thinking* is a broader category than *believing*. But adding the predicate ‘is true’ narrows the ‘thought’ category to something more closely associated with belief—although it still turns out to be too broad.

In particular, ‘is true’ draws our attention to various attitudes that are referred to in the literature as *regarding-as-true* attitudes. These attitudes include beliefs, conjectures, assumptions, suppositions, hypotheses, guesses, (propositional) imaginings, and many others, all of which involve regarding a proposition as true. For example, I can believe that p is true; assume that p is true; and guess that p is true. Each of these attitudes therefore have in common that they are ways of *thinking* a proposition *as true*. They are, to use a technical notion introduced by J. David Velleman, forms of *acceptance* (bearing in mind that we can accept a proposition without believing it, as when we accept that the moon is made of blue cheese, say, for the sake of argument).¹

So, this much at least seems true: believing belongs to a particular category of thought that involves regarding a proposition as true—a category we are calling *acceptances*. Nevertheless, we must also observe that believing is not like other forms of acceptance. To guess, to assume, to hypothesise is not to believe; believing is somehow a distinct way of accepting a proposition. Still, our initial assumption—that believing *just is* thinking true—cannot help us with this distinction; it brings us so far, but not far enough. To distinguish beliefs from other forms of acceptance, then, we need to explore different

¹ Velleman’s discussion of acceptances appears in his influential paper ‘On the Aim of Belief’ (2000b). In many ways, this thesis and the literature discussed throughout are reactions to (and sometimes defences of) the themes discussed in Velleman’s seminal paper. ‘On the Aim of Belief’ takes centre stage in Part I.

avenues; and this is the task that forms the central thread of this thesis: I want to know what beliefs are, which involves, specifically, knowing how to distinguish them from other forms of acceptance.

As far as I'm aware, the first philosopher to address this type of question was David Hume in his *A Treatise of Human Nature*. Hume aims to 'discover more fully the nature of belief', insofar as *that* nature distinguishes beliefs from other ways of 'conceiving' ideas (Book I, Part III, §7). Now, while Hume's own solution to this problem is ultimately unsatisfactory,² versions of his question continue to be a mainstay in philosophy of mind. As such, I follow the tradition and refer to the first problem of this thesis—that of distinguishing beliefs from other forms of acceptance—as Hume's Problem.³ In these terms, my aim is to explore various proposals for answering Hume's Problem, with the aim of finally offering my own solution. To achieve this end, I inquire into how to distinguish beliefs from other acceptances according to certain necessary and sufficient conditions.

In addition to this first question, I am also interested in two further problems concerning beliefs. First, I want to know whether we can provide a plausible account of why we, as believing beings, have beliefs. Or, to put things differently, why we have come to hold attitudes that satisfy the conditions for believing. Second, I want to say something, as is standard when providing a theory of belief, about the relation between belief and truth; and in particular I want to say something about *doxastic correctness*, the principle that it is correct to believe that p if and only if p . The first of these additional concerns I introduce because generally it is not adequately (if at all) addressed by philosophers proposing

² Hume wants to suggest that the difference between believing and other ways of conceiving ideas is a matter of degree. We conceive of our beliefs with an 'additional force and vivacity' by comparison to which our other thoughts are 'weak and languid' (Book I, Part III, §7). However, this doesn't seem right. When we think about, for example, guessing or conjecturing, these attitudes strike us as of a different *kind* to belief—not just as weaker or lesser versions of the same kind.

³ For example, see: Armstrong (1973) and more recently Van Leeuwen (2009).

theories of belief. Often, in the relevant literature, a lot of thought is put into delineating the necessary and sufficient conditions for belief, but little into why we have come to hold attitudes that satisfy those conditions. Nonetheless, knowing why we have beliefs is both interesting and crucially important. Interesting because, if we can provide an explanation, we will be better informed about the nature of our attitudes; and important because failing to make sense of why we have beliefs could be a reason in itself to reconsider our preferred theory of belief.

The second problem, concerning the principle of doxastic correctness, I discuss because of the important and commonly acknowledged relation between belief and true. In this respect, the question is: How should we interpret the seeming platitude that true beliefs are correct and false beliefs are incorrect? As we shall see, many take doxastic correctness to indicate something substantial about the relation between belief and truth. Some suggest that truth is, in this sense, the necessary aim (or *telos*) involved in believing (the teleologists, see Part I); while others claim that truth is the necessary *norm* involved in believing (the normativists, see Part II)—I call these readings of doxastic correctness *substantive*, because they take the principle to suggest an essential teleological or normative relation between belief and truth. In contrast to the substantive reading is (what I call) the *thin* reading. On the thin reading, doxastic correctness does not suggest an essential teleological or normative relation between belief and truth, but is instead interpreted as a pervasive yet *non-essential* social construct, such that members of our linguistic community use ‘true belief’ synonymously with ‘correct belief’ and likewise for ‘false belief’ and ‘incorrect belief’.⁴ Thus, there are, broadly speaking, two distinct ways to interpret doxastic correctness, which I discuss in more detail as we proceed. Ultimately, I agree

⁴ I have in mind for the thin reading of doxastic correctness something like Kathrin Glüer and Åsa Wikforss’ (2009, pp. 35-37) claim that classifying true beliefs as correct and false beliefs as incorrect amounts to a non-normative *categorising* of beliefs, similar to categorising items into tables and non-tables. But I also add some input from David Papineau (1999, 2013) who urges that doxastic correctness is merely a social convention. For further discussion, see Part III, Section 13.

with the thin reading, and reject the substantive reading (for my extended discussion see Part III, Section 13).

All things considered, my thesis addresses three related questions:

- (i) What are beliefs? (Hume's Problem)
- (ii) Why do we have them?
- (iii) How should we interpret doxastic correctness?

The first question forms the central thread, according to which my thesis is structured. The second and third questions are crucial considerations that I keep in mind as I analyse the plausibility of various theories of belief, including my own.

Thesis Structure

In the following, I explore answers to the above questions, with a particular focus on three kinds of theories of belief: teleological theories (Part I), normative theories (Part II), and functional theories (Part III)—hence the title of the thesis. I argue that the teleological and normative theories fail, in various respects, to provide plausible answers to the above questions, and that a functional theory, which I develop and defend, is to be preferred.⁵

In Part I, I focus on the teleological theories. The shared idea behind the teleological approach is that beliefs are subject to a unique aim, such that an acceptance subject to that particular aim is a belief. The two most common suggestions are that beliefs aim at *truth* (as per Bernard Williams' now famous remark⁶) and that beliefs aim at *knowledge*. Thus, on a teleological account, Hume's Problem is solved according to the proposed aim of belief. Acceptances that are subject to the aim are beliefs, and acceptances that are not subject to the aim are not beliefs.

⁵ For a useful edited edition, focusing in particular on teleological and normative theories of belief, see Chan (2013).

⁶ It is 'characteristic of beliefs that they aim at truth' (Williams 1970).

Doxastic correctness, furthermore, is given a substantive reading. Believers necessarily aim to believe the truth (or to know), so true beliefs are correct insofar as they are successful beliefs (or items of knowledge), while false beliefs are incorrect because they fail to satisfy their goal. Correctness here represents the successful satisfaction of a goal, and incorrectness represents failure. It might be wondered at this point whether this means doxastic correctness is a contingent principle on the teleological account, as the aim involved in believing could be otherwise—it is not. The point the teleologists want to make is that the aim of belief (whether truth or knowledge) is essential to the concept *belief*. If we were to talk about a different aim, then we simply would not be talking about the concept *belief*.

Finally, the second question—why we have beliefs—is glossed over by the teleologists. Some allusions to evolutionary processes are offered by a couple of teleologists. For instance, Asbjørn Steglich-Petersen (2006), whose work I discuss in detail later, writes in passing that the mechanisms that help realise belief's aim are 'presumably naturally selected' (p. 510). However, while I agree with a broadly evolutionary approach to understanding why we have beliefs, just mentioning natural selection in passing is not a satisfactory answer. As such, I do not attend to discuss how the teleologists might answer the second question in any detail, but rather provide my own evolutionary account (see Part III) of why we have beliefs; and I suggest only that, where this question is concerned, the teleologists have some work to do.

In assessing a teleological solution to Hume's Problem and a teleological account of doxastic correctness, I argue overall that the teleologists fail, for various reasons, to provide a plausible theory of belief, and that because of this, they also fail to provide a plausible account of doxastic correctness. That is to say, if belief is not in fact a teleological notion, then we cannot account for doxastic correctness in teleological terms. All things considered, I suggest, we should reject teleological theories of belief.

In Part II, I turn my attention to normative theories of belief. The thought behind normative approaches is that beliefs are subject to a unique normative requirement, such that an acceptance is a belief if and only if it is governed by this norm. As with the teleological theories, truth and knowledge are the most common suggestions. In other words, rather than a truth- or knowledge-aim, there is a truth- or knowledge-norm that is essential to belief. However, unlike the teleological theories, the normative account is not purely descriptive, precisely because it invokes normative entities.

For the normativists, Hume's Problem is answered according to the truth- or knowledge-norm. While there is nothing wrong (or defective), for instance, about assuming or hypothesising something that is not true (or that we don't know), the thought is that something *is* wrong with believing falsehoods. So, a truth or knowledge norm appears to govern beliefs, while the same is not true of other forms of acceptance. In this way, the normativists also make a lot out of doxastic correctness, because it appears to be a principle specifically of belief and not of other acceptances. If I assume something for the sake of argument, even if that thing is false, then it doesn't make sense to say my assumption is necessarily incorrect (possibly I know it is false but assume it anyway); but, if I believe something false, then it does make sense to say my belief is incorrect. *Correctness*, for most normativists, is a normative concept, and so doxastic correctness describes a normative truth about belief—that there is a requirement (to be later specified) to believe the truth and avoid believing falsehoods. On the normative account, then, doxastic correctness is given a substantive reading, such that the relation between belief and truth is essentially a normative one.

Of the theories I discuss, the normativists are least forthcoming with any account—in answer to the second question—of why we possess beliefs. Of why, that is, we possess a form of acceptance that is essentially subject to a truth- or knowledge-norm. The main argument in favour of this position is the norm's explanatory power. It helps to explain

what beliefs are, where, according to the normativists, a descriptive account fails; and it gives a plausible account of doxastic correctness, thus explaining why we take true beliefs to be correct and false beliefs incorrect. However, where these norms come from or why they have developed is unclear. I won't take too much time, in this respect, to analyse how the normativists might account for why we are subject to the belief norms, but will only mention that their account is, like the teleologists, lacking in this respect.

All things considered, I argue against a normative theory of belief. For various reasons, the proposal that beliefs are essentially normative is problematic; and, furthermore, I can see no good reason for accepting that correctness (and thus doxastic correctness) is a normative concept. As such, as with teleological theories of belief, I reject normative theories of belief.

In Part III, I begin to propose my own theory of belief. To answer Hume's problem, I suggest two theses that are necessary and jointly sufficient for counting an acceptance as a belief. The first thesis I take and develop from the *traditional motivational role* theory of belief. This is, broadly, the idea that beliefs (alongside desires) have the unique potential to cause and rationalise (*to motivate*) action. The second thesis I call the *fundamentality thesis*. This is the idea that beliefs occupy a fundamental role in our mental framework, such that one cannot possess other acceptances if one does not have beliefs about what one is accepting. In this sense, I argue, beliefs *inform the content* of other acceptances. Together, these theses enable us to distinguish beliefs from other forms of acceptance, and thus amount to my theory of belief.

To support this theory, I then provide an answer to the second question—why we have beliefs—thereby taking us further than previous theories of belief have. To do so, I turn to the literature in philosophy of biology. I suggest that we interpret the motivational and fundamental roles that beliefs occupy as *biological functions*. The idea is that, by interpreting beliefs as functional devices (as we do hearts, hands, and eyes etc.), we can gain an insight

into why we possess beliefs—they have developed via the processes of natural selection, roughly, because of their ability to perform their functions.

There are, however, two dominant theories of biological functions in the literature: etiological and systemic. Due to this, I dedicate Sections 8-11 to arguing in favour of a systemic interpretation of functions, which I then use to provide a functional analysis of my theory of belief.

To provide this analysis, I consider the motivational and fundamentality theses in turn, and thus assess how they can be interpreted as functional statements. Ultimately, my analysis will be partial, because a full systemic analysis of a functional device requires a complete account of all the interacting mechanisms and physical components that instantiate the device to be specified. This is simply not something I can do here, as it requires a detailed empirical investigation into the physical processes that enter into belief formation and maintenance—i.e., it requires the joint efforts of neuroscientists, cognitive scientists, and biologists. However, I can provide enough of an analysis to make progress toward explaining why we have beliefs.

With a systemic account of beliefs in place, I then turn my attention to the principle of doxastic correctness. From the functional perspective I offer, what can we make of the principle that true beliefs are always correct, and false beliefs always incorrect? As I have already mentioned, I support a *thin* reading of doxastic correctness, such that the principle is understood as a social construct, and not as indicative of any teleological or normative relation between belief and truth. On this understanding, it's not clear that a theory of belief is required to say anything about doxastic correctness. However, we should be willing to say something about why this social construct has developed—and as we shall see, the motivational thesis, in particular, goes some way to doing just that. In addition, in Section 13, I provide some reasons independent of the motivational thesis for

explaining why we should accept a thin reading of doxastic correctness, and I attempt to soften the blow of rejecting a substantive reading with a few further considerations.

Finally, in Part IV, with my theory of belief complete, I address one more important issue. Theories of belief often focus on providing conditions for outright belief and disbelief, however there is at least one other doxastic attitude: suspended belief (suspension). My question in this final part of the thesis concerns what the previously discussed theories of belief have to say about suspension. The assumption is that theories of belief, if complete, should be able to provide an account of all doxastic attitudes, including suspension. I argue that attempts to develop the teleological and normative theories of belief to account for suspension fail, but that with some modifications the functional theory of belief that I propose succeeds.

To conclude my thesis, I hold that a functional theory of belief—in which the motivational and fundamental roles of belief are interpreted as systemic functions—provides an answer to the three central questions of this thesis; and I suggest that the teleological and normative alternatives fail in these respects. My theory tells us what beliefs are, why we have them, and grants us a thin reading of doxastic correctness. Furthermore, when we extend the domain of a theory of belief to cover additional doxastic attitudes, and in particular suspension, my functional theory again wins out over the alternatives. Thus, all things considered, we should accept a functional theory of belief.

PART I
AIMS

Part I: Aims

1. Teleological Theories of Belief

The two most common teleological theories of belief hold that beliefs aim at truth and that beliefs aim at knowledge. Of these two hypotheses, the truth-aim is favoured amongst teleologists. Even those who prefer the knowledge-aim, such as Conor McHugh (2011), also accept that beliefs aim at truth. This is because they take the truth-aim to be ‘derivative from’ the knowledge-aim, granting that knowledge entails truth.⁷

For this reason, besides a discussion of McHugh’s teleological account in Section 2.1.2., most of my attention is on the truth-aim hypothesis. So, what exactly does it mean to say that beliefs aim at the truth?

Broadly speaking, the truth-aim thesis can be written as the following:

Truth-Aim: An acceptance φ is a belief if and only if φ aims at the truth.⁸

This thesis informs us that only acceptances formed subject to the truth-aim are beliefs, because being subject to the truth-aim is both a necessary and sufficient condition for belief. But we still need to know how the notion of *aiming at the truth* is unpacked.

According to the teleologists, there are two distinct ways that acceptances can aim at the truth: through our *conscious intention* to believe the truth, and through *subconscious processes* that regulate for truth. So, for example, Velleman writes that one way beliefs form is when ‘a person *intentionally* aims a belief at the truth, by forming it in an act of judgement’ (p.

⁷ For instance, McHugh (2011) writes about the knowledge-aim: ‘the suggestion is not that belief does not aim at truth, but that the truth-aim is derivative from the more fundamental aim of belief: knowledge’ (p. 382).

⁸ Theorists who commit to versions of this thesis, to be specified in more detail below, include: McHugh (2011, 2012), Sosa (2003, 2008, 2010), Steglich-Petersen (2006, 2009, 2017), and Velleman (2000). Since writing his seminal paper on the aim of belief, however, Velleman has switched allegiance to normativism about belief (see Shah and Velleman 2005)—I discuss normative theories of belief in Part II. Also, as already noted, McHugh thinks the more fundamental aim of belief is knowledge.

252, italics added). And Steglich-Petersen (2006) agrees that in the context of judgement (or in the context of deliberation, as I will say), ‘what one normally means by “intending to form a belief” is intending to accept a proposition if and only if that proposition is true’ (p. 511).⁹ Thus, the process of aiming at truth is made less abstract by the teleologists, by associating it, in the context of doxastic deliberation, with our *intention* to believe the truth. But of course many of our beliefs are not formed with our explicit intention to form beliefs, so the teleologists make the further appeal to subconscious processes that regulate beliefs for truth. For instance, Velleman writes:

A person can also aim cognitions at the truth without necessarily framing intentions about them. Suppose that one part of the person—call it a cognitive system—regulates some of his cognitions in ways designed¹⁰ to ensure that they are true, by forming, revising, and extinguishing them in response to evidence and argument (p. 253).

And Steglich-Petersen again echoes his sentiment:

Many, in fact most, of people’s beliefs are formed through subconscious processes of perception and inference which are not in any interesting sense controlled by the intentions of the subjects who have them. We thus need to construe truth-regulation in a wider sense so as to encompass non-intentional regulation (p. 502).

So, the truth-aim is also associated with subconscious processes that regulate for truth.

As such, we may restate the truth-aim hypothesis more precisely this time, as follows:

Truth-Aim*: An acceptance φ is a belief if and only if either:

⁹ For reasons discussed presently, Steglich-Petersen’s claim that the truth of a proposition is *sufficient* for us to intend to believe it is problematic. However, for now the important thing to note is the teleologists’ appeal to our intentions to elucidate the truth-aim.

¹⁰ By *designed* Velleman means *has been naturally selected for*. I discuss the teleologists’ appeal to natural selection later in this section. However, in Part III, Section 10.1, I also discuss in detail how speaking about natural selection in terms of design is problematic.

(a) a subject forms φ (in deliberation) with the intention to hold φ with content p only if p ,

or,

(b) a subject forms φ due to subconscious processes that regulate φ for content p only if p .

From (a), we can see that all beliefs formed in deliberation are done so according to the subject's explicit intention to hold that belief only if it is true. As Velleman writes, you entertain 'a question of the form " p or not p ?", wanting to accept whichever disjunct is true...' (p. 252)—you intend to believe that p only if p , or to disbelieve that p only if not- p .

Notice, also, that the truth of a proposition is necessary, but not sufficient, for us to intend to believe it. This formulation thus diverges from Steglich-Petersen's initial statement (cited above) that intending to form a belief involves 'intending to accept a proposition if and only if that proposition is true'. However, the divergence is justified, given that the sufficiency condition is clearly flawed. If, that is, the truth of a proposition were sufficient for us to intend to believe it, then we would have the intention to believe all truths. Yet clearly we do not have this intention—there are many truths that most of us do not care about at all, such as the number of atoms in the universe. And moreover, we are finite beings with a finite capacity for believing, yet there are infinitely many truths out there, so we cannot hope to grasp them all, even if we wanted to (Sullivan-Bissett and Noordhof 2013, p. 454). Thus, I prefer to formulate and discuss the truth-aim (at the intentional level) as the more plausible intention to believe only, but not all, truths. And, besides a brief reoccurrence of the problem of the sufficiency condition in Section 2.1.1., this remains the thesis I focus on throughout.

As for condition (b), this accounts for the many beliefs that, as the teleologists acknowledge, we form without explicitly intending to do so. When we are not in the process of deliberation, but are just going about our daily business, we form many beliefs, such as perceptual beliefs about our environment. Condition (b) captures these attitudes *as* beliefs (that aim at the truth) by proposing that they are regulated for truth by subconscious processes that are sensitive only to truth.¹¹ Again, it's important to note here that beliefs are said to be regulated for truth, but not that they regulated for *all* true propositions, since the processes involved clearly do not regulate for all truths. This is particularly clear when we reflect on the fact that our brains are not constantly (subconsciously) calculating the logical consequences of each one of our beliefs. Given our limited capacity, as finite beings, for inference, presumably this kind of constant subconscious regulation for truth would be mentally debilitating.

That covers the basics of the truth-aim (although we will consider some modifications to the initial hypothesis as we move along): beliefs, so defined, are essentially teleological, and among other acceptances, they are the only ones to aim at truth in this sense. Thus, Hume's Problem is solved according to these conditions. However, before getting critical of the teleological account, we still need to say a bit about the other two concerns of this thesis: why, according to the teleologists, we have beliefs; and what account the teleologists give of doxastic correctness.

About why we have beliefs, as I mentioned earlier, the teleologists do not have much to say. In the Introduction, I quoted Steglich-Petersen as saying that the processes involved in belief formation, those that regulate for truth, were 'presumably naturally selected'. But simply appealing to natural selection is not enough. We want to know more about how

¹¹ The thesis that subconscious belief forming processes are sensitive only to truth turns out to be false (as a counterexample, consider the processes of wishful thinking)—a point that is central to a dilemma Nishi Shah (2003) puts to the teleologists. However, I set aside this worry for now, to be discussed later in Section 1.2.

and why natural selection led to us having beliefs. In this respect, Velleman goes a little further with his appeal to ‘design’ by natural selection, by talking about how beliefs have the function to be true: they are ‘metaphorically speaking, designed to be true’ (Velleman 2010, p. 253, fn. 18). However, this raises further unanswered questions about what Velleman means by the *function* of belief, and what the relation between functions and natural selection is.¹² So, as I said earlier, the teleologists have more to say if they wish to explain why we have beliefs in evolutionary terms.

For my part, I agree that an appeal to natural selection is the correct move when we are trying to understand why we have beliefs, whatever the theory of belief we might eventually adopt. But if we are going to make this appeal, we need to be clear about what we mean. In this sense, the teleologists are lacking: they profess evolutionary commitments, but skimp on the details. I thus won’t say anything more about the teleologists’ appeal to natural selection; but instead try to show in Part III, concerning my own theory of belief, that an evolutionary account can be given.

As for the nature of doxastic correctness, true beliefs are correct on the teleological account because they are the successful satisfaction of either our intention to believe the truth or of the subconscious mechanisms that regulate for truth. And by the same token, false beliefs are incorrect because they are unsuccessful in these same respects—to have false beliefs, something must have gone wrong in the process of belief formation. For some reason, when we believe falsehoods, we fail to satisfy our intention to believe the truth (maybe we have insufficient or misleading evidence), or our beliefs have not been regulated properly (maybe our belief forming mechanisms are defective). On this interpretation, doxastic correctness is thus given a substantive reading, because there is

¹² To be fair to Velleman, his aim is explicitly not to naturalise the process of belief formation, but is instead to elucidate the thesis that beliefs aim at the truth (thus distinguishing them from other forms of acceptance). Nonetheless, for a complete account, these questions need to be addressed. This is what I intend to do for my account in Part III, Section [12].

an essential teleological relation between belief and truth. This, therefore, means that if the teleologist's account of belief is wrong, then so must be their account of doxastic correctness.

That said, I now begin to discuss and propose various problems with construing belief as a teleological concept. I begin by looking at whether being subject to the truth-aim is sufficient for distinguishing beliefs from other forms of acceptance.

1.1. The Sufficiency of the Truth-Aim

Assuming for now that the teleologists are correct—that beliefs aim at truth—is it really the case that the truth-aim distinguishes beliefs from other acceptances? In this section, I argue that it is not. In particular, I focus on how these forms of acceptance—*guessing* and (propositional) *imagining*—can also aim, in the relevant sense, at truth.

Of guessing and imagining, most obviously, guessing aims at the truth. This is a fact David Owens (2003) draws our attention to. To use his example, suppose you are a contestant on a quiz show and the host asks whether the population of the Earth is over 7.5 billion.¹³ You don't know the answer for sure, but if you guess correctly you will win £1 million. So, in the context of wanting to win the money, you aim to guess the truth, because that is the only way for you to win the money—you cannot guess falsely or not at all. In other words, you intend to guess that p only if p . As such, at the intentional level, guesses aim at the truth just as beliefs do.

One objection to this point is that guessing is not really an acceptance at all, but just a kind of speech act. In this sense, we can say *I guess that p* without intending to do so only if p is true. Of course, this is true, we can say we guess things without intending them to

¹³ I've increased the population from Owen's example. In 2017, guessing whether the population is over 7 billion is too obvious.

be true, but we can also say we believe things we don't take to be true. This doesn't mean that there is not a genuine attitude *guessing* that does not aim at the truth, just as it does not mean beliefs do not aim at truth. And in fact, there is something more to guessing than just performing lip service. For instance, we can imagine telling someone who commits to an absurd guess, such as that the population of the planet is less than 1000, that they cannot really be serious—that they are not *really* guessing. Moreover, if they insist that they are making a genuine guess, we can imagine pushing them by asking whether they really think what they are guessing is true. This indicates that the *guessing* concept we are using is not merely a speech act, but has more to it: it is a genuine kind of acceptance that a subject intends to hold only if it is true.

However, despite all of this, we need not rely on guessing alone to show that the truth-aim is not sufficient for distinguishing beliefs. Imagining can also, I argue, aim at the truth, and there is no dispute about whether imaginings are really acceptances.

To see how imagining can aim at the truth, in the sense that we can intend to imagine that p only if p , we first need to say a few things about the nature of imagining. In the literature, there is an important distinction between *imagistic* and *propositional* imagining. Imagistic imagining, as the name suggests, involves imagining with mental images. For instance, when you imagine that there is a monkey on your office desk, you may form a kind of image in your mind of a monkey on your desk. Propositional imagining, on the other hand, includes components that are not imagistic. For example, you might imagine that your neighbour is angry about you stealing his parking space. Sure, in this case, you may form a mental image of, say, your neighbour gritting his teeth and shaking his fist; but there must be something about your imagining that makes it one of your neighbour *angry*

and not just *pretending to be angry*. This, many agree, is the imaginings non-imagistic propositional content.¹⁴

So, propositional imagining can be thought of as a form of acceptance. It involves regarding-as-true certain propositions for the sake of imagining. To clarify, you imagine it as true that your neighbour is angry; and this is what distinguishes your imagining from one of your neighbour pretending to be angry.¹⁵

Another point about imagining is that we can intend to imagine. We can, that is, consider what we want to imagine, form the relevant intention, and go ahead and imagine it. For example, you can imagine your neighbour is angry, and then form various intentions about how your imagining is going to unfold—you may imagine him first as angry, then beginning to get sad, then beginning to form intentions about what he's going to do about the unidentified car in his parking space. This way of intending to imagine is what Kendall Walton (1990, pp. 13-16) calls *deliberate imagining*.

In this way, propositional imaginings are acceptances, in the relevant sense, and they can be subject to our intentions. We may, then, begin to wonder whether we can intend to imagine only that which is true, and I think we can. To begin to see this, consider that we can imagine things that we take to be true. For instance, you may be sat in the office after long hours of staring at a computer screen, and decide to imagine feeling comfortable at home, on your sofa, later that evening. And you can imagine this despite the fact that you also believe that you will be at home later, feeling comfortable on your sofa. So, we are capable of imagining things that we take to be true. Given this assumption, we can thus plausibly conceive of an individual who intends *only* to imagine the truth: he only, say,

¹⁴ See, for instance, Kendall Walton's discussion of imagistic and propositional imagining in his *Mimesis as Make-Believe* (1990, pp. 11-21).

¹⁵ Some, such as Amy Kind (2001), argue that all imagining is essentially imagistic. However, this is an extreme view. Even if images cannot be dispensed with in imaginings, accepting that they include propositional content is extremely beneficial for explanatory reasons. Thus, I will continue to adopt the more moderate view that imagining often (although maybe not always) involves propositional content.

intends to imagine his neighbour angry when he thinks he really is angry, and only imagines going to the comfort of his home when he thinks this is likely to happen. At the very least, therefore, having the intention to only imagine the truth is a conceptual possibility. Thus, if correct, imagining at the intentional level can aim at truth in just the same way as belief, meaning that condition (a) in the truth-aim hypothesis is not sufficient for distinguishing beliefs from other acceptances.

However, there is a possible objection to this line of reasoning. One might object that in these cases, in which an individual intends to have imaginings with true propositional content, the attitude just collapses into belief. In other words, intending to imagine the truth just causes the attitude of ‘imagining’ to be a belief, and not really an instance of imagining at all. So, when you imagine *and* believe your neighbour is angry, your imagining may involve mental images of your neighbour gritting his teeth etc., but the relevant attitude that makes the imagining one of him being angry, is a belief. Thus, imagining with the intention to imagine the truth is impossible—to do so is just to intend to believe. To this, I want to say a couple of things. First, my intuition is that in these cases imaginings and beliefs remain separate attitudes, even when we intend for the content of our imaginings to be true. To imagine in this sense, and to believe, are just two separate attitudes. This is because being part of the imagining is what makes the attitude an instance of imagining, and the fact that we may also believe a proposition that is part of an episode of imagining is irrelevant to whether we can also imagine it. This is an intuition also shared by others discussing imagining, thus Walton (1990) writes: ‘...imagining something is entirely compatible with knowing it to be true, (p. 13). And Neil Sinhababu (2012) agrees: ‘Imagining what I know to be true, after all, isn’t problematic’ (p. 156). Thus, I do not think intending to imagine the truth causes imagining to collapse into belief. We can imagine and believe the same thing at the same time. However, I do not want this matter to rest on an intuition alone, because we can also show that intending to imagine the truth

does not necessarily cause the attitude to collapse into belief. This is apparent when we consider that an individual is able to intend to imagine the truth, and form an instance of imagining, while suspending belief about (and thus *not* believing) what he imagines.

Consider a case similar to that of the quiz case, discussed above, involving guessing. Only this time, a prize is awarded for imagining. In particular, you are hooked up to a machine that can detect what you imagine, and can display the propositional content of your imaginings on a screen (in English). You are also told that to win £1 million you need to imagine either:

q: There is intelligent life in the Andromeda Galaxy.

not-q: There is not intelligent life in the Andromeda Galaxy.

Regardless of which is true, to win you just have to imagine one of them. The only way to lose is not to imagine either *q* or *not-q*. However, prior to the competition you have decided, as a personal commitment, to only ever intend to imagine the truth. So, wanting to win the prize, but also not wanting to break your personal commitment, you go ahead and form the imagining that you think is likely to be true. This way, you aim to satisfy both your intention to imagine the truth and to win the prize. As it happens, you go ahead and imagine *q* (because you read some Wikipedia entries on extra-terrestrial life and think *q* is more likely than *not-q*), and you win the prize for imagining.

Notice, however, that by intending to imagine the truth (and thereby going ahead and imagining *q*), your imagining clearly does not collapse into belief. Rather, you may very well suspend belief towards *q* or *not-q*, while still intending to imagine the truth. That is, you can suspend belief about *q*, because, say, you do not think the current evidence is sufficient to support an outright belief in *q*, while at the same time intending to imagine *q* only if *q*, because you think that, once the evidence does come in, *q* will be proved true. In this way, you can intend to imagine *q* only if *q*, and thus imagine *q*, while *not* forming

an outright belief in q . Thus, intending to imagine a proposition only if that proposition is true does not cause the resulting attitude to collapse into a belief—the two attitudes come apart.

As such, at the intentional level, both guessing and imagining (when construed appropriately), are counterexamples to condition (a) of the truth-aim as a sufficiency condition for distinguishing beliefs from other forms of acceptance. Even if we accept that we do intend to believe only the truth, we also intend to guess only the truth, and may decide even to imagine only the truth. However, so far the discussion has only focused on condition (a), our intention to believe the truth. We still may wonder whether condition (b) fares any better as a sufficiency condition. That is, maybe only beliefs are subconsciously regulated for truth.

To show why I think this is also incorrect, I again focus on imagining—only this time on instances of imagining that form subconsciously, without an explicit intention to imagine anything.¹⁶ First, then, we need to know what subconsciously forming imaginings amounts to. Consider again that you're at work, and that you've been staring at a computer screen for hours. This time, however, without intending to do so (you're trying to focus on your work!) you find yourself imagining that you are at home, feeling comfortable on your sofa. This imagining appears not thanks to your intention, but thanks to some subconscious mechanisms that, for whatever reason, cause you to imagine; and episodes of imagining like this can continue completely autonomous of our intentions, as, for instance, when you next find yourself also imagining getting in bed, closing your eyes, and thinking about how the duvet feels. This kind of imagining, which is independent of our

¹⁶ I won't focus again on guessing, because I'm not sure what it would mean to subconsciously guess. Possibly this would be some kind of conceptual confusion, because to guess seems to be to perform some kind of intentional mental act. Nonetheless, the example I discuss concerning imagining should be sufficient to make my point.

intentions, Walton (1990) calls *spontaneous imagining*—in contrast to deliberate imagining, that we were discussing above (pp. 13-16).

Hence, instances of imagining can occur despite us not intending to imagine. And given that this is the case, we can begin to create examples, similar to the one above, such that an individual has episodes of imagining that appear spontaneously, but that are regulated only for truth. In such a case, we just need to stipulate that the subconscious mechanisms involved in the spontaneous production of imaginings are truth regulated, rather than focus on the individual's explicit intention to imagine only the truth.

To be more specific, suppose that you are prone to spontaneous imagining, but that after banging your head you find that your spontaneous imaginings are only about things that you think more likely to be true, such as that you will be at home later today. Does this mean that your imaginings are in fact just beliefs, because they are regulated for truth? Of course, you may believe *as well* what you are imagining, but I think we should be cautious about making the assumption that imaginings regulated for truth are just beliefs. As previously stated, in the case of intentional imagining, the intuition supports the idea that intending to imagine the truth does not cause our imaginings to collapse into beliefs, and I think the same is true of imaginings that are regulated for truth. Nonetheless, we again do not need to let this rest on an intuition—based on similar examples as above, involving suspended belief and imagining, we can see that imaginings regulated for truth and beliefs are separate attitudes.

This time, think about some of the reasons why you find yourself spontaneously imagining (bearing in mind that after you bumped your head your spontaneous imaginings appear to be regulated for truth). You may, for instance, find yourself imagining being at home after work, because this brings you comfort; or you may, say, find yourself imagining things that you generally find interesting. Even though your imaginings are now

regulated for truth, these reasons for spontaneous imagining still apply. Thus, given that (we are assuming) one of your interests is extra-terrestrial life, you may find yourself imagining that intelligent life exists outside of the solar system. However, because your imaginings are now regulated for truth, you never find yourself spontaneously imagining that there is *not* intelligent life outside of the solar system. This is because, on reflection, you think the possibility of there being intelligent life (besides us) is greater than there not being. So, when spontaneously imagining, your imaginings are regulated for p only if p . However, just as with the intentional case, this does not mean that you necessarily also *believe* that there is intelligent life outside of the solar system: you may suspend belief about whether there is or not. As such, regulation for imaginings with true content does not mean that the resulting attitude formed is actually a belief.

Now, in objection to this case, one might say that what's really going on here is not that your imaginings are regulated for truth, but that they are regulated for what is more likely to be true. But I think if we say this, we also need to say that beliefs are not really regulated for truth but for what is more likely. For, we regularly form outright beliefs about things we are not certain about; and which, therefore, we just find more likely. For example, do you believe that your house is as you left it, and that it has not burned down—are you absolutely certain? So, this objection does not get off the ground. If we are willing to accept that beliefs are regulated for truth, as the teleologists are, we must also accept that in virtue of being regulated for truth, beliefs are formed according to subjective likelihood (and not because of absolute certainty); and the same can then be said of imaginings that are regulated for truth.

To summarise this section, I have argued that neither disjunct of the truth-aim hypothesis is sufficient for distinguishing beliefs from other forms of acceptance. This means that the teleological thesis is false. Even if beliefs aim, in the relevant sense, at the truth, this

aim does not distinguish beliefs from other forms of acceptance. In particular, guessing and imagining may aim at the truth in just the same way as the teleologists claim that beliefs do.

In the following section, I put the claim that the truth-aim is sufficient for distinguishing beliefs to one side, and I turn our attention to the necessity condition. I ask: do beliefs necessarily aim at the truth?

1.2. The Necessity of the Truth-Aim

As discussed, aiming at the truth is, on the teleological account, an essential feature of belief. All beliefs aim at the truth—but is this really the case? There are a number of reasons to doubt that this is so. In this section, I discuss these reasons. First, I focus on the intentional level, before I turn my attention to the subintentional level. Each time I ask whether beliefs really do aim at the truth in the way specified by the truth-aim hypothesis.

To begin, we may wonder whether we can intend to believe falsehoods at all. A quick survey of the literature on *intention* shows us that a common assumption is that: ‘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 [that action]’ (Frankish 2007, p. 534). So, to use Keith Frankish’s example, we cannot intend to pick a red card from a pack that is face down. And, similarly, the argument in the case of belief would be that we cannot intend to form false beliefs, because we have no control over whether we will succeed in doing so. We cannot, that is, go ahead and intentionally form beliefs that we take to be false. However, this is not quite right. Even if we accept the assumption about intentions—that we cannot intend to perform some act the success of which we have no control over—we may still form the intention to believe a falsehood. This is because we do have control over whether we succeed in forming false beliefs. We

may adopt strategies such as hypnosis or taking certain drugs that induce in us false beliefs. I could, for instance, take a drug knowing that every time I take it, within moments of ingesting it, I come to believe that there is a monkey on my office desk (even though there is not). So, I see no problem with saying that we can intend to believe falsely.

Nonetheless, the teleologists will reply that the point is not that we cannot intend to believe falsely, but that we cannot bring about a false belief due to this intention. The strategies just mentioned must always involve some kind of *indirect* method for believing falsely, and as such do not represent genuine cases of intending to believe falsely *and* bringing a false belief about in the relevant sense. In the moment the false belief is formed, even if I am hypnotised or in a drug induced state, I will form it with the intention to believe truly, and will thus only be able to do so because at that moment I take the belief to be true.

My question, however, is what work the notion of ‘indirect’ belief formation is doing here. If ‘indirectly’ is a temporal claim such that we do not satisfy our intentions to believe falsely instantly, then of course the teleologists are right, these methods are indirect. But this isn’t an obstacle to intending to believe falsely and thereby bringing false beliefs about. Most of our intentions we do not satisfy instantly, including any intentions we might have to believe the truth. When, for instance, I decide to learn some truths about a particular period of history, I need to take some action, such as going to the library, to learn these truths. So, the time taken between intending to form a false belief and forming that belief cannot be what’s at stake here.

Instead, if the teleologists mean by ‘indirectly’ that these strategies necessarily involve having to forget our intention to believe falsely, when the false belief is formed, then they have a more plausible claim. It does, it seems, appear to be a conceptual impossibility to take a belief to be false while at the same time believing it to be true. Thus, even, for

instance, in a drug induced state, I cannot believe that there is a monkey on my desk unless I take it to be true that there is a monkey on my desk, despite my prior intention to hold that belief only if it is false. Nevertheless, the fact that we cannot take a proposition to be false and simultaneously believe that same proposition does not entail that we cannot have an active intention to believe falsely right up until the point of forming a false belief. Our intention to believe falsely may be instrumental in the production of a false belief even though once the intention is satisfied we take what we previously thought to be false to now be true. This is at least a conceptual possibility that the teleologists deny, yet it is clearly demonstrated in Jonathan Bennett's (1990) race of believers he calls the Credamites. A Credamite is able to *will* himself to believe something he takes to be false, just because he wants to hold that belief. Bennett tells the story better than I can:

Credam is a community each of whose members can be immediately induced to acquire beliefs. It doesn't happen often, because they don't often think: 'I don't believe that P, but it would be good if I did'. Still, such thoughts come to them occasionally, and on some of those occasions the person succumbs to temptation and wills himself to have the desired belief. (Sometimes he merely wants to be the centre of attention and to amuse others. Someone who has no skill as an actor can instantly start to behave exactly as though he believed that P, by coming to believe that P. It is fun to watch it happen.) When a Credamite gets a belief in this way, he forgets that this is how he came by it. The belief is always one that he has entertained and has thought to have some evidence in its favour; though in the past he has rated the counter-evidence more highly, he could sanely have inclined the other way. When he wills himself to believe, that is what happens: he wills himself to find the other side more probable. After succeeding, he forgets that he willed himself to do it (Bennett 1990, p. 93).

The behaviour of the Credamites in bringing about false beliefs does not strike us as conceptually incoherent. Thus, having the desire to believe falsely, forming the intention to believe falsely, and bringing that intention to fruition is a conceptual possibility, despite what the teleologists claim.

That leaves us with subconscious belief formation to consider. And it is this disjunct of the truth-aim thesis, as a necessary condition for belief formation, which has caused the most trouble for the teleologists in the literature. Even those willing to accept that in conscious deliberation we necessarily intend to believe the truth, have found it difficult to accept that subconscious belief formation is also *only* regulated for truth.

The worry comes to light when we think of beliefs that are formed, for instance, by the processes of wishful thinking (call them wishful beliefs). These are beliefs that are influenced by our desires, preferences, or biases, even though we don't realise at the time that we hold those beliefs for non-truth-conducive (or non-evidential) reasons. At the intentional level, we hold wishful beliefs because we take them to be true; but at the subconscious level they are not influenced by evidential reasons. Thus, wishful beliefs are beliefs that at the subconscious level are not only regulated for truth, and there is a problem with the teleologist's thesis. This problem is made apparent in the dilemma Nishi Shah (2003) presents to the teleologists and that he aptly names *the teleologist's dilemma*.

On the one hand, we need a theory of belief that accounts for the exclusive interest we pay to evidential reasons in conscious deliberation. Or, as Shah puts it, we need a theory that can explain *transparency*, the thesis that the question *whether to believe that p* is answered by, and only by, answering the question *whether p is true* (p. 447).¹⁷ And on the other hand,

¹⁷ This version of transparency is similar to, although should not be confused with, an earlier version of transparency that was used as a solution to the problem of how we can know our own beliefs (see Evans 1982 and Moran 1988; 2001, pp. 60-65). I discuss the earlier version of transparency briefly in Section 5.2.1.

we need a theory of belief that accounts for the influence of non-evidential factors in subconscious belief formation, thus granting wishful beliefs the status of bona fide beliefs. Thus, if we accept the truth-aim thesis as it stands, we can account for transparency, because beliefs are regulated (at both the intentional and subconscious levels) only for truth. But the consequence of this is that it rules out cases of non-truth regulated beliefs—such as wishful beliefs—as genuine beliefs. However, if we alter the truth-aim thesis to allow for wishful beliefs, such that beliefs are not only regulated for truth, then we lose our explanation of why only evidential reasons factor into conscious deliberation. So the teleologists have a problem—what are they to make of Shah’s dilemma, and in particular of the observation that at the subconscious level beliefs are not only regulated for truth? For one, these considerations have caused Velleman to abandon his teleological account of belief in favour of Shah’s suggestion: that beliefs are essentially normative.¹⁸ But others have not been so easily deterred, and attempts have been made to defend the truth-aim hypothesis. In response, Steglich-Petersen (2006) argues that there is not in fact one concept of belief in use, but two. The first concept is the one we have been discussing, but only at the intentional level: in conscious deliberation we intend to believe that p only if p . But at the subconscious level, granting that non-evidential reasons can factor into belief formation, the belief concept that applies is merely *weakly* regulated for truth. So, while in conscious deliberation we only allow what appear to us to be evidential reasons to influence what we believe, we also accept that non-evidential reasons may affect belief formation at the subconscious level.

¹⁸ See Shah and Velleman (2005). I discuss normative theories of belief in Part II.

Now, as I mentioned, these two concepts are related, and according to Steglich-Petersen we can only understand the subconscious belief concept in relation to the intentional belief concept (which he calls the primary concept). He writes:

So the primary concept of believing *p* is that of accepting *p* with the aim [or intention] of doing so only if *p* is true. Even if [the beliefs] fail to achieve that goal and are merely weakly responsive to truth, cognitive processes can count as instances of beliefs in virtue of being *brought about by someone who has this aim*, with the important qualification of being *at least to some degree conducive to this aim* (2006, p. 515, italics added).

There are two important things to note here, which I have put in italics. The first is that the individual who holds weakly-truth-regulated beliefs must also have the primary aim of intending to believe only the truth. The second is that weak regulation for truth must be at least partly conducive to satisfying the individual's primary aim to believe the truth. However, on reflection, there are problems with each of these points: in the first case, a belief can be regulated for truth even if the individual holding that belief does not intend to believe the truth (he may not intend to believe anything at all); in the second case, at least some subconsciously formed beliefs are not conducive to the believing the truth, and so are not even weakly regulated for truth. On Steglich-Petersen's account, then, these attitudes would not be counted as beliefs, yet in our ordinary conceptual framework we do count them as beliefs.

Let's begin with the first problem, which is slightly tangential to our concern in this section, but is nevertheless a central part of the proposed relation between the two belief concepts. Beliefs, despite what Steglich-Petersen says, can be regulated for truth in the absence of any intention to believe the truth—i.e. in the absence of applying the primary belief concept. Charles Côté-Bouchard (2016) provides us with some examples. Consider for instance that you are reading a newspaper and don't want to see last night's football

score, because you recorded the game and plan to watch it later. However, by accidentally turning to the wrong page, you inadvertently see the result, and can't help but form a belief about it. In this case, the processes that lead to you forming a belief about the game's score regulate for truth (assuming that the paper is reliable) despite you actively intending to avoid holding a true belief. So, even if our beliefs are regulated for truth at the subconscious level, we do not also necessarily need to intend to have true beliefs. Similarly, we may be completely disinterested in whether we have a particular true belief, yet form it anyway because of subconscious processes that regulate for truth. For instance, you may not care at all about certain historical facts, such as the life of Napoleon Bonparte, but on hearing on the radio that he was born on the island of Corsica, you come to believe it. Despite your disinterest, again, your subconscious belief forming mechanisms may still regulate for truth.¹⁹ Thus, both of these kinds of cases demonstrate that there is a breakdown in the relation Steglich-Petersen proposes between the primary belief concept (at the intentional level) and the subconscious belief concept. Our beliefs may be regulated for truth even if we are disinterested in or actively intend to avoid the truth of a particular proposition.

The second problem with Steglich-Petersen's proposal, which is more central to our immediate concern, is that there can be beliefs that are not even weakly regulated for truth, despite the fact (assuming now that it is a fact) that we necessarily intend to believe only the truth. That is, our intention to believe the truth tells us nothing about the regulation of our beliefs at the subconscious level—on first person reflection, this information is just not available to us. It is conceptually possible, that is, to imagine a race of believers who necessarily intend to hold beliefs only if they are true, but who never in fact form true beliefs because the subconscious mechanisms involved in the production of their

¹⁹ See Côté-Bouchard (2016) for these examples. I have altered the historical case but the point remains the same.

beliefs are systematically regulated to ensure that they are false. To these believers, their beliefs will always appear true to them in the first person (as ours do to us), yet they will almost never in fact be conducive to their primary intention to believe the truth, and will nearly always be false.²⁰ So nothing about necessarily intending to believe the truth should cause us to think that our beliefs at the subconscious level are regulated for truth, not even weakly.

Moreover, this problem can be emphasised using Steglich-Petersen's (2006) own analogy between *belief* and *concealment* (pp. 511-515). According to Steglich-Petersen, concealing an item necessarily involves intending to place it where it cannot be found (at least by the person from whom you intend to conceal it). So if you want to conceal some chocolates from your sister, and if your act is to count as a genuine act of concealing, you must place the chocolates somewhere you think your sister will not find them (p. 512). If you intentionally place the chocolates right in front of your sister, say, and claim to be concealing them, we may think either that you are joking or that you are mistaken about what it means to conceal something. However, we may still perform a genuine act of concealing, from the first person perspective, even if the item we intend to conceal is in fact not put in a place that is conducive to our aim—that is, if we don't hide the item well. Thus, just like with belief, there is a primary and secondary concept of belief. The primary concept involves, in the first person, necessarily intending to put an item only where it cannot be found (as believing involves intending to believe that *p* only if *p*); but the secondary concept allows that we may not conceal an item well, and thus that it might be found (as our beliefs may turn out to be false). This secondary concept, then, just like the subconscious belief concept, is only *weakly* regulated for its end.

²⁰ I say *almost never* and *nearly always* because even beliefs regulated for falsity will sometimes be true by accident, so to speak.

Now, what makes the secondary concealment concept weakly regulated for its end is analogous to the subconscious belief concept: the person concealing the item must *have the primary intention to put the item where it cannot be found* and the item must be placed somewhere that is *at least to some degree conducive to satisfying this intention*. But, just as our intention to believe the truth does not tell us anything about how our subconscious regulates for beliefs, our intention to conceal an item does not tell us anything about whether the item is really put somewhere that is even to some degree conducive to concealing it. For instance, it might turn out that when you are concealing the chocolates from your sister, she is sat secretly watching you the whole time. From your perspective, you are still concealing the chocolates, yet where you placed them was not even partly conducive to your aim: in no sense did you in fact put them somewhere that your sister cannot find them. Thus, just as with belief, the relation between the primary concept of concealment (that you intend to place an item only where it cannot be found), and the secondary concept (that the act of concealing must be at least to some degree conducive to satisfying your intention) breaks down. There is no reason to suppose that our intention to put an item only where it cannot be found means that in our attempt to do so our act is weakly regulated for satisfying that end; and there is no reason to suppose that our intention to believe the truth means that our beliefs are at the subconscious level regulated even weakly for truth. As I have said, it is entirely plausible that a race of believers could exist that have the intention to believe that p only if p , while at the same time having subconscious belief forming mechanisms that do not regulate for truth. Thus, rather than serve to elucidate his two concepts of belief, Steglich-Petersen's analogy serves to highlight the disparity between the two concepts he proposes.

With this said, while I have argued that subconscious beliefs may (as a conceptual possibility) not be regulated for truth, I do not think that *our* beliefs, as humans, are all systematically not regulated for truth. Many of our beliefs do appear to be regulated, to

some degree, for truth. Obvious examples include immediate perceptual beliefs in ordinary circumstances. Nonetheless, a more moderate suggestion, which I think is true, is that *some* of our beliefs are not even weakly regulated for truth. Here are two reasons to suppose this is the case.

First, we have no problem bringing to mind people who are systematically subject to the processes of wishful thinking. These people, no matter what evidence they receive, and no matter what the testimonies of others are, are never deterred from holding some of their wishful beliefs. For instance, imagine the father who is absolutely convinced that his son is going to be a professional football player; or your aunt who buys ten lottery tickets every week certain that she will get lucky soon enough. Such people are among us, and they have wishful beliefs that are systematically not regulated for truth.

Second, empirical evidence from evolutionary psychology suggests that we have evolved to have beliefs about certain things that are not regulated for truth. If true, this means that all of us are subject to forming some beliefs that are systematically not regulated for truth, because that is how our belief forming mechanisms are structured. In the literature, these beliefs are referred to as *adaptive misbeliefs* and the general idea is this: believing falsehoods in certain areas has given us an evolutionary advantage, such that we have adapted to hold false beliefs in those areas. One of the least controversial types of adaptive misbeliefs that we are thought to have are *positive illusions*. These include conflated beliefs about our own personal abilities, appearance, and achievements, and of those close to us, such as our children or spouse. For instance, when surveyed, most people profess to be better than average at almost everything that is subjective and socially desirable (Myers 2002). And the reason for these positive illusions, it is suggested, is that they are beneficial for our mental health (Taylor and Brown 1988) and our physical health (McKay and Dennett 2009). If the evidence is correct, then, and we have adaptive misbeliefs, we again have

reason to suspect that not all beliefs—both as a conceptual possibility and as a possibility in this world—are regulated, even weakly, for truth.²¹

Given these considerations, therefore, Steglich-Petersen's appeal to a second belief concept that involves, at the subconscious level, weak regulation for truth, is just as problematic as the initial thesis that beliefs are only regulated for truth. For a start, the link between what Steglich-Petersen calls the primary belief concept (which operates at the intentional level) and the secondary concept (at the subconscious level) breaks down. He gives us no reasons to think that because we intend to believe only the truth (assuming that we do) that our subconscious mechanisms regulate even weakly for truth. Furthermore, there are cases of attitudes that we would ordinarily call beliefs that are systematically not regulated for truth; this is not just a conceptual possibility but is a physical possibility that is supported by empirical evidence.

In summary, the teleologist's thesis that beliefs necessarily aim at the truth should be abandoned. It is a conceptual possibility to intend to believe falsely and as a result of this intention form a false belief—this is demonstrated by the Credamites. Furthermore, we can intend to believe falsely and via certain strategies, such as hypnosis or drugs, cause ourselves to form false beliefs. Additionally, the claim that beliefs are subconsciously regulated for truth is also problematic. In its initial form, the thesis suffers from obvious counterexamples, such as wishful beliefs; and in its modified form, such that subconscious belief formation only weakly regulates for truth, it remains flawed: beliefs can be formed that are not even weakly regulated for truth.

On the weight of these arguments, therefore, the truth-aim thesis fails to provide a plausible answer to Hume's Problem. Beliefs are not acceptances that aim at the truth in

²¹ For a survey of the literature on adaptive misbeliefs see McKay and Dennett (2009). I speak again about adaptive misbeliefs in Part III, Section 13.

the relevant sense; and if they were, this would not be sufficient for distinguishing them from other acceptances anyway—bearing in mind that guesses and imaginings can aim at the truth. Moreover, given that beliefs do not necessarily aim at the truth in the way the teleologists suggest, the truth-aim is also unable to account for doxastic correctness. If some beliefs do not aim at the truth, then true beliefs cannot be correct *because* they are the goal involved in believing.

With these objections raised, I turn my attention in the following section to perhaps the most discussed objection to teleological theories of belief. Specifically, I focus on a difficulty raised by Owens (2003) that has since come to be known as the *exclusivity objection* (Steglich-Petersen 2009).

2. The Exclusivity Objection

The exclusivity objection gives central importance to the idea that, when we consciously form a belief, we focus exclusively on considerations pertaining to the truth of that belief's content, i.e. on evidential reasons—call this the exclusivity thesis. On first sight, the exclusivity thesis appears to have a ready *explanation* in the form of the truth-aim thesis. If our intention is to believe p only if p , then we are only going to take interest in evidential reasons when forming beliefs. However, the exclusivity thesis is precisely the reason, according to Owens, that we *should not* accept a teleological theory of belief.

As Owens (2003) argues, the fact (assuming for now that it is a fact) that we cannot form beliefs in conscious deliberation for anything other than evidential reasons indicates that beliefs are not really teleological at all. Typically, Owens observes, *aim-motivated* behaviours can be weighed against other aim-motivated behaviours before an action is taken, and a compromise can be reached about how to proceed—yet, by hypothesis, the exclusivity thesis rules out any chance of weighing the aim of belief with our other aims. Hence, beliefs are not in fact teleological.

In this section, I begin by elucidating the exclusivity objection in more detail. I then discuss two responses the teleologists offer to the objection, and explain why each is problematic. The first is Steglich-Petersen's (2009) insistence that the aim of belief can be weighed against other aims, against which I invoke Ema Sullivan-Bissett and Paul Noordhof's (2013) response. The second is Conor McHugh's (2012, 2013) rejection of the exclusivity thesis in favour of a form of pragmatism about belief, which I argue requires McHugh to accept two conflicting theses.

2.1. Explaining the Exclusivity Objection

This following argument precisely captures the exclusivity objection:

(P1) Behaviour φ is aim-motivated only if deliberation about whether to φ can be weighed against other aim-motivated behaviours.

(P2) Deliberation about whether to form beliefs cannot be weighed against other aim-motivated behaviours.

(C) Therefore, deliberation about whether to form beliefs is not an aim-motivated behaviour (i.e. is not teleological).

Supporting premise one is the principle that all aim-motivated behaviours can be weighed against other aim-motivated behaviours. We bear witness to this principle when we take any behaviour that strikes us as aim-motivated, and consider the possibility of compromising the pursuit of that aim because it comes into conflict with another of our aims. For example, suppose that your aim is to be a good saxophone player. And you decide that the best way to achieve your aim is to spend twelve hours a day practicing. However, your wife takes no joy in your persistent saxophone playing, although she is happy for you to play for eight hours a day, while she is at work. So, rather than play twelve hours a day you decide to respect your wife's wishes and play just eight. Why?

Because you also have the aim of maintaining a happy family life, and by limiting your saxophone playing you can continue to pursue your aim to be a good saxophone player *while also* pursuing your aim to maintain a happy family life. In other words, you can consider both of your aims, and reach a compromise about how to pursue both of them—you can weigh them against each other. The point is that this observation generalises to support premise one: all aim-motivated behaviours must have the potential to be weighed against other aim-motivated behaviours.

Supporting premise two is the exclusivity thesis. If conscious belief formation permits only evidential influences, then the aim of belief cannot be weighed against other aims. Specifically, we cannot (even in part) form beliefs in full consciousness for non-evidential (i.e. pragmatic) reasons. To demonstrate the truth of the exclusivity thesis examples are usually given that involve our inability to believe a falsehood even when there are strong pragmatic motivations for doing so. For instance, we cannot form beliefs in propositions that we take to be false simply because we are offered a large financial reward.

Assuming, therefore, that the reasons supporting each of the premises are persuasive, we must conclude that beliefs are not essentially teleological.

In response to the exclusivity objection, the teleologists can of course attack either premise. To attack premise one, they can deny that all aim-motivated behaviours can be weighed. To do so, examples may be given of behaviours that we generally accept to be aim-motivated, but that cannot be weighed against other aim-motivated behaviours. So far, however, this stand has not been made by the teleologists—they are happy to assume (at least for the sake of argument) that premise one is true. As such, I too will not consider it any further.

Towards premise two the teleologists are more critical. Steglich-Petersen argues that the exclusivity thesis does not entail that forming beliefs cannot be weighed against our other

aims; and McHugh denies exclusivity altogether. Thus, on both of their accounts, premise two is false, though for different reasons. I now consider in turn each of their defences of the teleological approach to understanding beliefs.²²

2.1.1. Can the Aim of Belief be Weighed?

In order to understand Steglich-Petersen's (2009) reply to the exclusivity objection, we first need to know how he interprets the idea of an aim being weighed. In particular, he observes that there are two senses in which an aim can be weighed.

The first involves the weighing of *mutually compatible* aims. This kind of weighing is possible when an individual has two or more aims, and a course of action is available that permits the pursuit of each aim. One example of mutually compatible aims being weighed is the one discussed above, of your aim to become a good saxophone player and your aim to keep a happy family life. These aims are mutually compatible because you do not have to sacrifice one in favour of the other: you can reach a compromise about how to satisfy both aims and pursue them simultaneously. Another example of mutually compatible aims being weighed, and the one Steglich-Petersen offers, involves having the aim to clean the dishes while also having the aim to minimise your environmental impact. To satisfy both aims, you may decide to limit your hot water usage and to use environmentally friendly detergent (p. 401). In both of these examples, we see mutually compatible aims being weighed against each other.

The second sense of aims being weighed Steglich-Petersen describes involves *mutually incompatible* aims. This kind of weighing occurs when an individual has two or more aims

²² In discussion, Tony Booth mentioned a third reason that we might reject premise two: maybe the aim of belief can be weighed, but it just so happens that it always wins out over our other aims. However, I don't think this is a position anyone accepting the exclusivity thesis can agree with. To do so would make the fact that we believe only for evidential reasons contingent on the strength of those reasons, albeit that they are always stronger than our other reasons for believing. Yet exclusivity, if it is true, is supposed to be a conceptual and thus necessary truth about beliefs; such that even if our pragmatic reasons for believing do happen to be stronger than our evidential reasons, they still cannot influence belief formation.

that are in conflict, meaning that a single course of action is not available for the pursuit of both aims simultaneously. In this case, one or more of an individual's aims must be discarded so that another aim can be pursued. For instance, your aim not to drink alcohol during October and your aim to get drunk at your friend's birthday party on October 23rd are mutually incompatible—one must be discarded in favour of the other. The importance of mutually incompatible aims, according to Steglich-Petersen, is that they demonstrate generally that aims do not need to be compatible to interact, and thus do not have to be compatible to be weighed. Just like mutually compatible aims, mutually incompatible aims can be weighed in deliberation. The difference is the outcome: one of the aims must be discarded (pp. 401-402).

So what is the relevance of all this to the aim of belief? Steglich-Petersen proposes that while beliefs have no mutually compatible aims, they can be weighed against mutually incompatible aims. In his words:

My hypothesis is that while the aim of belief uniformly fails in being capable of the kind of weighing against other aims that results in a decision that combines the pursuit of both aims at once, the aim of belief is capable of the kind of weighing that results in discarding either of the aims because of mutual incompatibility (Steglich-Petersen 2009, p. 402).

If this is correct, then the second premise of the exclusivity objection is false—the aim of belief can in principle be weighed against other aim-motivated behaviours. All that is required is for other aims to come into conflict with our aim to believe the truth, such that we can choose, if we wish, to discard the aim of belief in favour of the other aim.

To demonstrate how this happens, Steglich-Petersen provides two kinds of examples. The first involves deciding whether to *take up* the aim of belief, and the second involves deciding whether to *give up* the aim of belief.

In the first instance, when deciding whether to take up the aim of belief, an individual may face various obstacles. For instance, the matter of how ‘practically feasible’ it is to enter into an inquiry concerning the truth of a proposition. To use Steglich-Petersen’s own example, we may be deciding whether to inquire into whether global warming is caused by human activity. To make this inquiry, however, would require a lot of resources and effort that we are unwilling to commit. Thus, instead of pursuing the inquiry, we decide never to take up the aim of believing the truth about whether humans cause global warming, because it conflicts with our other aims (namely, those to do with resources and effort). In this way, then, the aim of belief can be weighed against other (mutually incompatible) aims: the practicalities of pursuing a line of inquiry may influence you not to take up the aim of belief (p. 403).

In the second instance, an individual may decide to give up the aim of belief. This happens when ‘one comes to realise that the aim would be incompatible with, or detract from the feasibility of some other aim’ (p. 403). For example, Steglich-Petersen asks us to imagine a teacher who wants to know which student broke a window, but who decides not to continue her inquiry because of the unpleasant task of scolding the guilty student that will inevitably follow if she finds out the truth. In this case, the teacher initially has the aim to believe the truth, but upon realising the consequences of having the belief, decides to give up on her inquiry, thus abandoning the aim of belief. So again, the conclusion Steglich-Petersen draws is that the aim of belief can be weighed in the sense of mutually incompatible aims. We can discard our aim to believe the truth in favour of some other aim.²³

²³ By proposing that the aim of belief can be weighed in these ways, it’s important to note that Steglich-Petersen is not promoting a form of pragmatism. Rather, his claim is that the interaction between our aim to believe the truth and our pragmatic aims is such that the latter does not *influence* the former, but instead requires us to abandon it. So, on Steglich-Petersen’s account, pragmatic considerations do not enter into belief formation, despite his claim that they can interact with the aim of belief.

Now, however, comes the problem. As Sullivan-Bissett and Noordhof (2013) argue, for Steglich-Petersen's position to hold, he must rely on an implausible formulation of the truth-aim thesis. Recall from Section 1 that the aim of belief, at the intentional level, is

to intend to believe that p only if p

and *not*

to intend to believe that p if and only if p .

As previously noted, this is because the second version of the aim makes the truth of a proposition sufficient for us to intend to believe it; but intending to believe all true propositions would require us 'to fill our minds up with loads and loads of useless, trivial truths', and it is clearly not an intention that we, as ordinary believers, have (Sullivan-Bissett and Noordhof 2013, p. 454).

The issue then is that for Steglich-Petersen's examples of taking up or giving up the aim of belief to count as genuine cases of belief's aim being weighed, he must be committed to the implausible second interpretation of the aim. This is because only the second interpretation, and not the first, requires us to enter into inquiries in pursuit of truths. To see this, reflect on the fact that the first interpretation (i.e. to intend to believe that p only if p) does not make the pursuit of truths a necessary part of the aim of belief. Instead, it suggests a conditional of the form *if you form a belief intend only for that belief to be true*. So, it does not commit us to seeking out truths—and concerning Steglich-Petersen's examples, it does not require us to form any beliefs about global warming or which student smashed the window. It only requires us to intend to believe the truth about these propositions should we in fact form a belief about them. In Sullivan-Bissett and Noordhof's own words: 'Since the truth aim does not require that we form beliefs, it does not go into the balance when we consider the question whether it is worthwhile to form a belief' (p. 455).

However, if Steglich-Petersen adopts the second interpretation of belief's aim, which includes the sufficiency condition, he has what he needs. Given that there is a truth about whether global warming is caused by human activity, and a truth about which student broke the window, the second interpretation would require us to begin an inquiry into these matters—and we may *then* weigh this aim with pragmatic considerations, such as practical limitations. Nevertheless, as stated, this version of belief's aim is completely implausible; so if Steglich-Petersen is depending on it—as it seems he must—his theory is in trouble.²⁴

For these reasons, I agree with Sullivan-Bissett and Noordhof that Steglich-Petersen's attempt to avoid the exclusivity objection fails. Instead of demonstrating that the aim of belief can be weighed against other aim-motivated behaviours, he shows that an implausible interpretation of belief's aim can be weighed against other aim-motivated behaviours. Thus, I see no reason, from what Steglich-Petersen says, to think that premise two of the exclusivity objection is false.

I now turn my attention to McHugh's alternative attempt to undermine the exclusivity objection by rejecting the exclusivity thesis and instead accepting a form of pragmatism.

2.1.2. Rejecting Exclusivity: McHugh's Pragmatism

Central to McHugh's (2012, 2013) response to the exclusivity objection is the debate between evidentialists and pragmatists about the reasons for which we can form beliefs in doxastic deliberation. In short, McHugh argues for a form of pragmatism that entails that the exclusivity thesis (and along with it evidentialism) is false: at times, in conscious deliberation, pragmatic considerations can (at least in part) be taken into account as

²⁴ On a more plausible interpretation, Steglich-Petersen is really talking about weighing the aim of *inquiry* (i.e. to discover truths) with pragmatic aims. However, even so, this doesn't help his case insofar as he is attempting to show that the aim of belief, and not inquiry, can be weighed against other aims.

reasons for belief. Thus, beliefs are not subject exclusively to evidential reasons and the aim of belief can, therefore, be weighed against pragmatic aims. If this is true, then premise two of the exclusivity objection is false.

In this section, however, I argue that the two theses McHugh adopts to make his argument go through are mutually exclusive. On the one hand, he wants to remain committed to the thesis that beliefs aim at truth (and also, as we will see in his case, knowledge); and on the other hand, he wants to accept that in certain cases of belief formation we have *discretion* over what to believe (where discretion permits us on occasion to take into account pragmatic concerns when forming a belief). The problem, or so I argue, is that if we accept the truth- or knowledge-aim, then we cannot have discretion over what to believe; and if we accept that we have discretion over what to believe, then our beliefs do not necessarily aim at truth or knowledge. So McHugh is faced with a dilemma: either accept that we do not have discretion over what to believe, and thus have no solution to the exclusivity problem, or accept that we do have discretion over what to believe, but at the expense of giving up a teleological theory of belief.

To see how this unfolds, we first need to have a clear view of McHugh's teleological commitments, which are slightly different from the truth-aim thesis that we have been discussing so far, given that he ultimately prefers a knowledge-aim thesis. And then we need to get to grips with his *discretion principle*: the thesis that we sometimes have discretion over what to believe. I discuss each these ideas in turn. Finally, I show why McHugh's knowledge-aim thesis, along with the truth-aim thesis, cannot be accepted simultaneously with the discretion principle.

2.1.2.1. McHugh's Teleological Thesis

In similar fashion to the truth-aim thesis, McHugh associates the aim of belief with the *regulation* of belief for a particular end—in his case, knowledge. He writes:

I will take it that a state aims, in the relevant sense, to satisfy condition *C*, when it is the state's nature to be regulated, either through conscious reasoning or through non-conscious cognition, by processes that have as an aim or function that *C* be satisfied (2011, p. 370).

On this analysis of an aim, then, beliefs aim at knowledge through regulative processes that operate in a person's explicit intention (in conscious reasoning) or in his subconscious cognition (in non-conscious cognition). Thus, as with the other teleologists, McHugh gives belief's aim a literal reading as opposed to a metaphorical reading.²⁵

In addition, McHugh also makes the aim of belief essential for an acceptance to count as a belief. He puts the point in terms of constitution: "The teleological conception thus takes the notion of an aim literally: there really is an aim... that is constitutive of belief" (p. 370). Thus, granting that constitution implies necessity, the claim that beliefs are constituted by a knowledge-aim implies that they are necessarily subject to that aim.²⁶ Any attitudes that do not aim at knowledge, on McHugh's account, are not beliefs. This is an important point when we come to see that using discretion over what to believe entails rejecting the idea that beliefs are necessarily subject to a knowledge- or truth-aim.²⁷

Finally, and this is where McHugh's teleological thesis really diverges from the truth-aim thesis, he takes the primary aim of belief to be knowledge. So, rather than merely aiming at truth, beliefs aim at knowledge. But what does this claim amount to? McHugh tells us that: "To believe *p* on the basis of evidence you regard as insufficient would be to have a

²⁵ The normativists claim that beliefs aim at truth in the metaphorical sense that they are governed by a norm such that the only appropriate thing to believe is the truth (i.e. a truth-norm). I discuss normative theories of belief in Part II.

²⁶ Although constitution implies necessity, philosophers usually mean something more when they say, for example, that concept *A* is constitutive of concept *B*. Typically, it means something like: we cannot *understand* or *grasp* concept *B* without understanding also concept *A*. So, to say that belief is constituted by an aim is to say that we do not understand what beliefs are unless we also understand that they are necessarily subject to that aim (cf. Boghossian 2003, p. 37).

²⁷ McHugh (2011) also accepts that being subject to a knowledge-aim is sufficient for an acceptance to count as a belief. Nonetheless, I omit discussion of his sufficiency condition here because it is irrelevant to the objection I want to raise.

belief that, by your own lights, doesn't amount to knowledge' (p. 382). This gives an important role to evidence in the knowledge aim. You must 'by your own lights' (i.e. subjectively) have enough evidence for your belief, if you form one, to amount to knowledge.²⁸ Also, McHugh adds:

You cannot, deliberately and in full awareness, form an outright belief in a proposition, if you regard your evidence for that proposition as not putting you in a position to know it (p. 383).

So, if you don't think your evidence is sufficient for knowing that p , you won't be able to form a belief that p . In aiming at knowledge, therefore, you must have sufficient subjective evidence for putting you in a position to know that p , otherwise you cannot form a belief that p . Now, given that knowledge entails truth, McHugh also accepts that beliefs aim at truth, in the sense we were discussing in previous sections. He writes that 'the suggestion is not that belief does not aim at truth, but that the truth-aim is derivative from the more fundamental aim of belief: knowledge'. Therefore, on McHugh's account, beliefs also aim truth in virtue of aiming at knowledge.

Given these points, we can formulate McHugh's knowledge-aim thesis as follows:

Knowledge-Aim: An acceptance φ is a belief only if φ is regulated for the truth of a proposition p and the subjective evidence a subject S has for p is sufficient for putting S in a position to know that p .²⁹

The problem with this thesis is still the exclusivity objection. If our beliefs are only regulated for truths which we have sufficient evidence to know, then we still would not take pragmatic considerations to be relevant to belief formation. As such, the knowledge-

²⁸ The evidence we require for forming beliefs is *subjectively* sufficient for knowledge because if sufficient *objective* evidence was required all beliefs would necessarily amount to knowledge, but this is obviously false.

²⁹ The knowledge-aim thesis can be broken down into a disjunctive thesis about how beliefs regulate via our intentions or via subconscious processes that regulate for knowledge, just as we saw with the truth-aim thesis. However, there is no need for such detail here.

aim cannot be weighed against pragmatic aims in belief formation. However, this is why McHugh introduces the discretion principle.

2.1.2.2. McHugh's Discretion Principle

To avoid the exclusivity objection, McHugh (2012, 2013) argues that in certain situations we have discretion over what to believe. These situations occur when we have *sufficient but non-compelling* evidence for a belief. To be specific, McHugh (2013) writes that 'your evidence for p can be such that you *can* believe p , but you can also withhold belief in p ' (p. 1122). Hence, the discretion principle can be stated as:

Discretion Principle: A subject S can have sufficient but non-compelling evidence for believing p , such that S can either believe that p or withhold belief in p .

If true, this principle means that when we have discretion over what to believe, pragmatic considerations can factor into whether we form a belief. Thus, exclusivity would be false, and we could reject premise two of the exclusivity objection. To support the discretion principle, McHugh offers a number of examples. Here I focus on just one, which should be enough to show how discretion works in practice.

Criminal Case

Several independent witnesses report that your friend has committed a terrible crime. Although it is possible that the witnesses are mistaken or lying, you do not think you would have this evidence if your friend were not guilty. However, you value your friendship, and think that believing in your friend's guilt will damage that friendship. This is particularly so if your friend turns out to be innocent.³⁰

In this case, McHugh claims that you have (according to the discretion principle) two options psychologically available to you. You can form the belief that your friend is guilty

³⁰ This example is paraphrased from McHugh (2013, p. 1124).

because you have sufficient evidence to do so; or you can withhold belief in your friend's guilt because you value your friendship. Given these possibilities, you can therefore *weigh* the pragmatic consideration (the concern for your friendship) into the balance when deciding whether to believe in, or withhold belief about, your friend's guilt. Thus, if examples such as this are telling, our aim to have beliefs that amount to knowledge can at least sometimes be weighed against our pragmatic aims—in particular, when we have sufficient but non-compelling evidence for a proposition and can thereby decide to either believe or withhold belief in that proposition.³¹

For the purpose of the objection I want to raise, the important point is that the discretion principle entails that beliefs are, at times, not *only* regulated for knowledge (or truth), but also for pragmatic concerns. On the surface, this is not a problem for McHugh, because the knowledge-aim does not rule out the possibility that beliefs are regulated for knowledge *and* other things. However, when we realise that enacting discretion requires us to reject the knowledge-aim, we see that McHugh has a problem.

2.1.2.3. The Conflict

To tease out the tension between accepting the knowledge-aim and the discretion principle, I'll begin with an analogy. Imagine that your aim is to buy a new car, and you have sufficient means of buying one; however, you decide not to buy the car because your wife says it's a waste of money. In this case, your decision not to buy the car is not

³¹ Sophie Archer (2015) argues that discretion is, in fact, impossible. She focuses on McHugh's claim concerning testimony. When you receive testimony from an informer, you can either decide to believe her because you take her to be reliable, or you can decide to withhold belief because you think she might be mistaken or lying (McHugh 2013, p. 1122). Archer points out that in such cases, we do not have discretion over whether to form an outright belief or withhold belief. Rather, there is a shift in evidence that determines which doxastic attitude we adopt. In the first case, your evidence is that your informer is reliable, which causes you to believe her; in the second case, your evidence is that your informer might be mistaken or lying, which causes you to disbelieve her. According to Archer, all apparent cases of discretion can be reinterpreted like this, so we should reject the discretion principle—there is no reason to bring pragmatic aims into the picture. I have sympathy with Archer's objection; however, I put it to one side because I want to raise a different objection. Namely, that if we do accept discretion, we must reject the knowledge- and truth-aim theses. For more arguments against the plausibility of discretion, see Kurt Sylvan (2016).

influenced by your initial aim (to buy the car), it is a rejection of it—specifically, in favour of your aim to keep your wife happy. Analogously, in the case of belief, if you have sufficient subjective evidence (sufficient means) to know a proposition, and yet for pragmatic reasons do not go ahead and believe it (instead withholding belief in it), your withheld belief is not influenced by your aim to hold only beliefs that amount to knowledge, it is a rejection of it. Why? Because you have sufficient means for satisfying the knowledge-aim, yet you decide not to form a belief. There can be no other explanation for you not believing that which you have sufficient evidence for knowing, other than that you gave up the knowledge-aim.

To bring my point to life consider again the criminal case. You can satisfy the knowledge-aim by believing in your friend's guilt, because by hypothesis you have what you would usually consider to be sufficient evidence for believing in his guilt (i.e. the testimonies of several independent witnesses); yet you reject this evidence and instead withhold belief because you don't want to risk ruining your friendship. Thus, by not believing in your friend's guilt, you are rejecting the knowledge-aim, because you have sufficient means of satisfying it (sufficient evidence), and you choose not to.

As such, there is an apparent conflict in accepting that beliefs aim at knowledge and that we have discretion over what to believe. To be more precise about this problem, it helps to focus again on the details of the knowledge-aim. For a belief to aim at knowledge, it must satisfy two conditions:

- (a) it must be regulated for the truth of a proposition p ; and,
- (b) it must be formed by a subject S who has sufficient subjective evidence to be in a position to know that p .

Hence, if the discretion principle entails that the knowledge-aim is false, as I have suggested, at least one of these conditions must be false when we enact discretion over what to believe. And the issue I want to raise lies with (a).

To see the problem, consider that in cases of discretion, we can withhold belief in a proposition despite having sufficient subjective evidence for the *truth* of that proposition—this is because we can withhold belief despite having sufficient subjective evidence to know that proposition, and knowledge entails truth. So, again in the criminal case, you have sufficient subjective evidence for believing that the proposition *my friend is guilty* is true, but instead you withhold belief towards that proposition. As such, the regulation of your belief for truth is replaced with the regulation of your belief for your pragmatic aim, which causes you to withhold belief. If this were not the case, and your pragmatic aim did not replace your aim to believe the truth, then you would just form the belief in your friend's guilt (you have sufficient evidence to do so!).

This is further reflected in the reasons you would give for not believing in your friend's guilt. You can imagine explaining to an outsider, who asks why you are withholding belief when the evidence suggests your friend is guilty, that you don't want to believe it (and so withhold belief) because you don't want to ruin your friendship. Your reason for withholding belief then is pragmatic, you have given up on believing what the evidence suggests.

Thus, accepting the discretion principle entails that there are cases of belief formation that do not satisfy condition (a) of the knowledge-aim: beliefs formed when we enact discretion are not regulated for truth. This, in turn, implies that the knowledge-aim is false if we do have discretion over what to believe, because regulation for truth is a necessary requirement for an acceptance to count as a belief. Furthermore, given that condition (a) of the knowledge-aim just is the truth-aim, discretion cases also undermine the truth-aim

thesis. That is, a theorist cannot appeal to discretion cases to avoid the exclusivity objection whether he accepts either the knowledge-aim or merely the truth-aim thesis.

This, then, leaves McHugh with a dilemma. Either the knowledge-aim or the discretion principle must go. If he accepts the discretion principle, then he cannot commit to the thesis that beliefs aim at knowledge. While if he accepts the knowledge-aim, he cannot appeal to discretion to avoid the exclusivity objection. Thus, McHugh has not shown that the exclusivity objection can be avoided by a teleological theory of belief.

At this point, one important worry may come to mind. In arguing for the discretion principle, McHugh appeals to cases of withheld belief, and argues that we can withhold beliefs for pragmatic reasons. As such, we may be tempted to think that the knowledge-aim does in fact still apply to beliefs even though it does not apply to withheld beliefs. We can enact discretion over whether to form a belief, while still aiming to believe only that which we can know. Thus, the knowledge-aim and the discretion principle can co-exist. However, for McHugh's defence against the exclusivity objection to go through, withheld belief must also count as a kind of belief in its own right (i.e. as a genuine doxastic attitude). If it did not, then McHugh's examples would not show that the aim of belief can be weighed with pragmatic aims, but instead that pragmatic aims can influence whether we form a non-doxastic attitude. This is a result that would bear no relevance to whether pragmatic aims can influence belief formation.³²

All things considered, then, McHugh's attempt to avoid the exclusivity objection fails because he does not show that we can accept a teleological theory of belief while also accepting that pragmatic considerations can influence belief formation.

³² I discuss withheld belief—or *suspended belief*, as I later refer to it—in detail in Part IV.

3. Summary of ‘Aims’

In Part I of this thesis, I laid out and discussed the core ideas associated with teleological theories of belief. The essential point the teleologists want to make is that beliefs are regulated, either via our intention or via subconscious processes, for a particular end—this end being, most commonly, truth. In particular, they want to say that being subject to these regulative processes is a necessary and sufficient requirement for an acceptance to count as a belief.

However, as we have seen, these basic notions of belief fail to address our three central questions. As I mention early on, the teleologists do not go into enough detail about why we have come to have attitudes that aim, in the suggested sense, at truth. Thus, they fail to answer our third question: why we have beliefs. But also, as was the main focus in Part I, they fail to give a plausible answer to Hume’s Problem and, relatedly, a plausible account of doxastic correctness.

When it comes to distinguishing beliefs from other forms of acceptance, the truth-aim is not sufficient—guessing also aims at the truth, and so can imaginings in certain contexts, given that the aim is associated with how our attitudes are regulated. Nor does the truth-aim appear to be a necessary condition for an acceptance to count as a belief. Bennett’s race of Credamites indicate that intending to believe the truth is not a necessary condition for *belief*; and we too may intend to believe falsely, as ordinary humans, and bring this about, at least by indirect means, such as through hypnosis or by taking drugs.

Furthermore, the failure of the teleologists to respond adequately to Owens’ exclusivity objection is another reason to doubt that a teleological conception of belief is plausible. In particular, Steglich-Petersen makes the truth-aim an implausible requirement in his attempt to avoid the objection, insofar as he makes it necessary for us to intend to believe all truths; and McHugh provides us with an argument for a pragmatic theory of belief,

which if we accept, entails that we have to give up the very theory that he sets out to defend; namely, that beliefs aim at knowledge.

For these reasons, I hold that the teleological thesis fails to answer Hume's Problem. We need a different account of how to distinguish beliefs from other forms of acceptance. But in addition, the same arguments provide reasons for rejecting the teleologist's account of doxastic correctness. For if beliefs do not aim at truth, in this teleological sense, then true beliefs cannot be correct *because* they aim at truth.

Therefore, given the considerations made in this Part of the thesis, the best move is plausibly to discard the teleological theory of belief, and instead focus on a different way of understanding beliefs. Hence, I now turn to consider normative theories of belief as a possible alternative.

PART II
NORMS

Part II: Norms

4. Normative Theories of Belief

Normative theories of belief have existed in the literature for some time; Griffiths (1962-63), for instance, writes that we pick beliefs out in the public domain as the attitudes that are ‘*appropriate to truth*’ (p. 182, italics original). In today’s terminology, this amounts to the claim that that our concept *belief* refers to an attitude governed by a truth-norm—that is, beliefs are appropriate to truth. More recently, normative theories of belief have become popular as an alternative to purely descriptive theories; such as the teleological theories we were discussing in Part I. The thought, for reasons to be discussed, is that a normative theory of belief can succeed where a purely descriptive theory fails.

Similarly to teleological theories, normative theories usually focus on truth or knowledge: the most common suggestions being that beliefs are subject to a truth- or knowledge-norm. Accepting again, however, that knowledge entails truth, the shared commitment of normativists is that beliefs are governed by a truth-norm, even if there is a more fundamental knowledge-norm. For this reason, as the case was with the truth-/knowledge-aim, most of my attention is on the truth-norm. The essential agreement of the normativists is thus:

Truth-Norm: An acceptance φ is a belief if and only if φ is governed by a truth-norm.³³

According to this thesis, beliefs are governed by a truth-norm, and other acceptances are not. This is because the truth-norm is both a necessary and sufficient condition for acceptances to count as beliefs. However, we may wonder why we should accept that

³³ Philosophers who agree that beliefs are governed by a truth-norm include: Boghossian (1989, 2003); Engel (2005, 2007, 2013a, b, c, 2017); Gibbard (2003, 2005); Shah (2003); Shah and Velleman (2005); Wedgwood (2002, 2007, 2013a, b); and Whiting (2010, 2013). Pascal Engel in particular, however, thinks that knowledge is the fundamental norm of belief (see Engel 2005). For a general survey of the literature on normative theories of belief see, McHugh and Whiting (2014).

beliefs are essentially normative, and what exactly it means to say that beliefs are subject to a truth-norm.

The precise nature of the truth-norm is a complex topic, and although the normativists agree that there *is* a truth-norm governing belief, there is little agreement on what the norm actually amounts to. Some interpret the norm as a *permission*, such that we may only believe the truth (Whiting 2010, 2013); others interpret it as a *requirement* such that we ought only to believe the truth (Boghossian 2003; Shah 2003; Shah and Velleman 2005); and others still interpret it as a requirement to believe all true propositions that we *consider*.³⁴ Each of these proposals must confront various problems, however I won't get into this issue here, for while there are many concerns surrounding how to formulate the truth-norm, none of them entail that *there is not* a plausible way to formulate it.³⁵ Instead, for the purposes of this thesis, I adopt perhaps the least problematic (and most common) formulation of the norm, which is to express it as a requirement.

Hence, my focus is on the following more precise account of the truth-norm thesis:

Truth-Norm*: An acceptance φ is a belief if and only if a subject S ought to hold φ with content p only if p .³⁶

As such, I talk throughout in terms of the truth-norm as a requirement, however as should be clear as we progress, the problems I discuss readily apply to normativism generally, and therefore to other possible formulations of the truth-norm.

As for why we should accept normativism, some theorists claim that beliefs must be normative because we cannot account for them solely in descriptive terms; and others

³⁴ This last proposal is offered by Ralph Wedgwood (see Bykvist & Hattiangadi 2007, p. 280).

³⁵ For a detailed discussion of the various problems facing the different formulations of the truth-norm see Bykvist and Hattiangadi (2007, 2013).

³⁶ This requirement tells us that we ought only to believe truths, but *not* that we ought to believe all truths. As theorists on both sides of the debate about normativism have noted (e.g. Bykvist and Hattiangadi 2007; and Boghossian 2003), we can hardly be required to believe all truths, given that *ought implies can*, and we can never hope to believe every truth out there.

take belief to be normative because of the normative significance they attribute to doxastic correctness.

Those who take the first approach do so as a response to the teleologist's dilemma, which we discussed earlier (see Section 1.2.). The general idea is as follows, given that there are instances of beliefs that are not regulated for truth, such as wishful beliefs and adaptive misbeliefs, we cannot say that beliefs are regulated only for truth. However, in conscious deliberation, we do only focus on evidential reasons for belief. So, to count non-truth regulated beliefs *as* beliefs, we can say that as believers, we accept that our beliefs are governed by a truth-norm. Thus, even attitudes that are not regulated for truth can count as beliefs if we also accept that they are only appropriate to hold if they are true. In this way, interpreting belief as a normative concept is seen as essential to solving (or at least part of the solution to) the teleologist's dilemma.³⁷

The other route to normativism, and perhaps the most common, is from a normative interpretation of doxastic correctness; and in particular, of *correctness*. From this perspective, *correctness* is seen as a normative concept in its own right. And for this reason, the fact that true beliefs are correct and false beliefs are incorrect entails that beliefs are only appropriate when true. In contrast to teleological theories of belief, then, these normative theories do not explain the correctness of belief as a goal that beliefs (or believers) necessarily aim to satisfy; rather, they simply take doxastic correctness to state a normative fact about beliefs, one that our beliefs are subject to regardless of any aim that we might have when we form beliefs.³⁸

³⁷ For a full account of this position see, Shah (2003) and Shah and Velleman (2005). Velleman changed his position from accepting a teleological theory of belief to instead adopting a normative position for this reason.

³⁸ Theorists who explicitly accept doxastic correctness as a reason for adopting normativism include: Boghossian (1989, 2003); Engel (2007, 2013a, c, 2017); Gibbard (2005); Wedgwood (2002, 2007, 2013b); Whiting (2010, 2013).

Important for my project, then, is that doxastic correctness is (in my terminology) given a substantive *normative* reading. The principle of doxastic correctness states an essential and important link between belief and truth, such that we cannot understand the concept *belief* unless we also understand this normative relation. This is true whichever approach we take to giving a normative theory of belief: whether we think beliefs must be normative because descriptive accounts are problematic, or whether we take correctness to be a normative concept in its own right, doxastic correctness states a normative fact about *belief*: specifically, that beliefs are appropriate when true and inappropriate when false. As such, we can deal with Hume's Problem and the nature of doxastic correctness simultaneously. In particular, if there is a problem with a normative account of belief, such that we cannot answer Hume's Problem with reference to the norm of belief, then there is also a problem with a normative reading of doxastic correctness.

Finally, this leaves one important concern left to answer—how can normativists explain why we have come to possess an attitude that is essentially normative? Notice that this is a different question to why we should think that beliefs *are* normative. It is a question about why we have beliefs at all. In regards to this question, as far as I am aware, the normativists are silent. Perhaps it is not considered important, or perhaps they think that showing beliefs must be normative is a kind of solution to this problem. However, as I have stated previously, the question is important in its own right, and showing that beliefs *must* be normative, if that is something the normativists achieve, tells us nothing interesting about why we possess beliefs.

One possible answer the normativists might be give is that by accepting that beliefs are governed by a truth-norm, we have given ourselves an evolutionary advantage, and so beliefs have the (biological) function of being true. They may then suggest that functions are essentially normative, and so claim that only true beliefs are appropriate, because only

true beliefs satisfy the function involved in believing. However, besides not knowing whether this is a route the normativists would want to take, it seems to me that it would be problematic; in particular, because I do not think functions are essentially normative (as I argue in Section 10).

Nonetheless, the main point is that the issue of why we have beliefs has not yet been addressed by the normativists, and so I cannot discuss how they might wish to answer this question in any detail. Instead, I'll just point out, as I did with the teleologists, that normativism about belief is lacking in this area.

With this basic outline of normativism given, I now turn my attention to perhaps the two most pressing problems facing normativism: *the no-guidance objection* and the *regress of rules problem*, which collectively I refer to as the *objections from normative impotence*.

5. Objections from Normative Impotence

In this section, I explore the consequences of the no-guidance objection and the regress of rules problem, each of which were posed in their original forms by Kathrin Glüer and Åsa Wikforss.³⁹ I discuss each of the objections in turn, with a particular focus on the plausibility of Pascal Engel's (2013a, c) response to the problems. I argue, in particular, that Engel's responses are ultimately unsuccessful.

5.1. The No-Guidance Objection

The essential idea behind the no-guidance objection is that norms typically provide guidance concerning the behaviour that they govern, but the truth-norm fails to provide guidance when we form beliefs. When getting to grips with the details of this argument,

³⁹ For the original statement of the no-guidance objection see Glüer and Wikforss (2009, §1.2), and for further discussion see Glüer and Wikforss (2010, 2013, 2015). For the regress of rules problem see Glüer and Wikforss (2009).

it helps to first understand some of the history of a similar and influential argument against normative theories of belief, which serves as a precursor to the no-guidance objection.

In particular, the thought that we cannot choose what to believe—i.e., the assumption that doxastic involuntarism is true—is often said to tell against any normative account of belief, since norms don't apply to actions or behaviours that we have no control over. To give an analogy, consider that you accept the following charity-norm: that you ought to give to charity. Due to accepting this norm, on occasion, you give to charity. The charity-norm prescribes something for you to do, and for that reason you go ahead and do it. However, notice that you don't have to give to charity: you can instead choose to defy the norm, even if you accept it, by not giving to charity. That option is available to you.

But now consider this reading-norm: that you ought to read twelve novels every day. Presumably, for ordinary humans, this is an impossible norm to satisfy. We just can't read fast enough to get through that many novels every day. Nonetheless, suppose that our friend tells us that we really ought to be obeying the reading-norm. In that situation, we would be in our rights to say to our friend something like: *what do you mean I ought to read twelve novels a day, I can't!* So, why is this a reasonable response? It is because norms must permit a certain amount of what Peter Railton calls (2003) *normative freedom*.⁴⁰ His idea is that for a norm to be genuinely normative, it must imply *both* that we can choose to act according to it *and* that we can choose to act against it. In other words, *ought implies can*, but also, *ought implies can not*. So, while in the charity-norm case you have normative freedom over whether to act according to the norm, in the reading-norm case you do not. Hence, the reading-norm is not genuinely normative.

Applying this reasoning to belief, we see that the truth-norm runs into some trouble. If doxastic involuntarism is true, because we cannot help but believe what we take to be true,

⁴⁰ For his full discussion see Railton (2003, pp. 322-51).

then we do not have the kind of control over our beliefs that is sufficient to enact normative freedom. That is, we cannot decide whether to act according to the truth-norm or against it, because we necessarily attempt to act according to it. So, the truth-norm breaks the *ought implies can not* rule. As such, doxastic involuntarism entails that beliefs cannot be governed by a truth-norm (and in fact cannot be normative).

Notable proponents of this line of reasoning include William Alston (1988) and Alvin Plantinga (1993). In particular, Alston argues that the kind of control we have over our beliefs is limited in such a way that it is insufficient to ground any kind of doxastic normativity (pp. 260-83); and Plantinga makes the same point in considering self-presenting propositions such as *I am sad* and *I am appeared to redly*:

If the fact is I am sad and I consider this proposition, then whether or not I accept it is simply not up to me; but then accepting this proposition cannot be a way in which I can fulfill my obligation to the truth, or, indeed, any obligation to try to bring about some state of affairs. Suppose I've just fallen off a cliff: could I be subject to an obligation to try to bring something about, which is such that I can better fulfill it by falling down rather than, say, by falling up or remaining suspended in midair? Hardly. (1993, p. 38)

If these points are to be believed, then doxastic involuntarism is problematic for the normativists.⁴¹ We are in our right to ask how there can be a norm that recommends forming true beliefs when we cannot avoid doing so (that is, when in a position to do so). Nonetheless, as we may expect, this is not the end of the story for the normativists. For, they have a ready reply to the objection from doxastic involuntarism.

⁴¹ As Plantinga (1993) notes, the objection from doxastic involuntarism should not be taken to generalise to all epistemic deontology. There can still be norms governing, for instance, inquiry, such as 'one ought to sufficiently *reflect* on one's evidence for a belief, since we have control over whether to perform such requirements (p. 24). Nonetheless, the point still holds for forming beliefs, which we seemingly do not have control over.

In response, the normativists argue that the objection rests on a misunderstanding of the nature of the truth-norm, and in spite of the fact that we cannot help but believe what we take to be true, we still *must* be committed to the norm in order to have beliefs. This is because, they say, the truth-norm is *constitutive* of belief, and if we understand constitutive norms properly, we see that doxastic involuntarism is not an obstacle to normativism.⁴²

The idea behind the constitutive role of the truth-norm is that we cannot have beliefs at all unless they are under the governance of the norm, because the norm is an essential part of what it means for an attitude to count as a belief. In this sense, the truth-norm is to *belief* as rules are to games—you must be committed to the rules in order to be playing the game. To be more specific, if you have a chessboard before you, and you move the chess pieces around at random, while having no idea what the rules of chess are, then you can hardly be said to be playing chess; and likewise, if you form beliefs with no respect for the fact that they ought to be true, then you are not really forming beliefs. Interpreted in this way—as constitutive—therefore, the fact that a rule is necessarily followed as a part of some act does not entail that the rule is not genuinely normative. To play games, we necessarily follow the rules; and to believe, we necessarily follow the rules, according to the normativists.⁴³

As such, the objection from doxastic involuntarism misses an important point: the truth-norm is a constitutive feature of belief. And instead of presenting a problem for the normativists, doxastic involuntarism is in fact explained by the truth-norm. Given that we necessarily commit to the norm when forming beliefs, we also necessarily only take into account evidential reasons for belief. If we did not, we would not be respecting the truth-norm, and thus would not be forming beliefs.

⁴² For example see, Boghossian (2003, p. 39), Shah (2002, p. 437), and Wedgwood (2002, p. 268).

⁴³ For further discussion and examples of constitutive norms see Railton (2003, ch. 10).

At this point, the objection from doxastic involuntarism loses its traction. Our inability to control our beliefs does not entail that beliefs are not normative. However, as we presently see, by continuing to focus on the relation between belief and truth, the no-guidance objection emerges as a more refined (and somewhat altered) version of the argument from doxastic involuntarism. According to Glüer and Wikforss (2009), the fixation on doxastic involuntarism was always ‘a bit of a red herring’ (p. 46). Yet the relation between belief and truth still leads to a problem for the normativists; that is, involving the ability of the norm to *guide* our belief formation.

In contrast to the objection from doxastic involuntarism, the problem is not simply that we cannot control what we believe; but is instead that the truth-norm provides no guidance when it comes to forming beliefs. Prescriptions, that is, typically provide guidance—they say what to do in certain circumstances; or, more abstractly, they say: ‘Do X when in C’. So, we want to know whether we can be guided by the truth-norm in this sense.

To understand whether the truth-norm can guide belief formation, we need to draw our attention to the difference between *subjective* and *objective* norms. For subjective norms we know that we are in C, and so can know when to perform X; while for objective norms we need not know that we are in C, and so need not know whether we are really in a position to satisfy the norm by performing X. However, despite this epistemic difference, both norms can provide guidance. For example, the charity-norm I mentioned earlier is a subjective norm. Usually, you can know whether you are giving to charity or not. You can choose a charity that you trust, sign up to their monthly donation program, and make regular donations to the charity. As such, you know that you are in circumstance C (a position to give to charity) and you can perform X (by giving to charity). Moreover, you do X (at least in part) *because* of the norm. In this way, the charity-norm provides guidance.

In contrast, the norm for maximising your profits on the stock market—*buy low, sell high!*—is objective.⁴⁴ With norms such as this, it is difficult to know that one is really selling high in relation to previous lows *especially* when the ‘high’ is understood as a *peak* in the share price. In this sense, whether we are in a position to ‘sell high’ is not ‘transparent’ to us (to use Boghossian’s term). Nevertheless, we can still be guided by the norm to buy low and sell high, even if epistemic problems prevent us from satisfying it.

Thus, whichever way we look at it, norms provide guidance—and this is the case even if we are wrong about being a circumstance C for satisfying a norm. But what is the relevance of this to the truth-norm? Well, the truth-norm, according to normativists such as Boghossian, is of the objective kind: we cannot always know whether we are obeying the norm given the information that may be available at the time of forming a belief; yet, the norm still governs belief formation, as a requirement we commit to satisfying when forming beliefs. As Boghossian (2003) writes: ‘The truth is what you ought to believe, whether or not you know how to go about it, and whether or not you know if you have attained it’ (p. 39).

However, this turns out to be exactly where the no-guidance objection takes hold. As we learn from Glüer and Wikforss (2009), the truth-norm is not really like other objective norms, despite Boghossian’s claim. While with other objective norms we can consider whether we are in C, and then look to the norm for guidance, and then perform the prescribed action X, with the truth-norm we are not able to look to it for guidance concerning what to believe. The problem is that ‘circumstance C, in this case, refers to the truth of a proposition’ (p. 44). So, the truth-norm suggests that when you are in a position to take *p* as true, you should go ahead and believe that *p*. But, when you are in that position, such that you take *p* to be true, you have already formed the belief that *p*.

⁴⁴ I take this example from Boghossian (2003, p. 45, fn. 6).

Our reasoning when forming beliefs does not go: *I take p to be true, the truth-norm says to believe true propositions, so now I'll believe that p* . As such, the truth-norm does not provide guidance—it tells us only that we ought to do what we, upon being in the relevant circumstance, have inevitably done.

Another way to think about this problem, which I think is helpful for visualising exactly what is going on, is to notice that there is no inferential step between *recognising that a proposition is true* and *believing that proposition*. To see this, consider the condition *if you recognise p to be true, then you ought to believe that p* . When the antecedent obtains, and you recognise that p is true, you do not then consider the normative ‘ought’ and submit to its requirements; rather, the consequent of believing that p occurs as a matter of logical necessity, i.e. when the antecedent condition obtains, you believe that p . This lack of an inferential step leaves no room for us to reflect on the truth-norm, and ask ourselves what guidance it provides; so, the norm does not provide guidance in the way typical of other prescriptive norms.

Viewed in this way, we can see more clearly the difference between the earlier objection to normativism, from doxastic involuntarism, and the more refined no-guidance objection. The latter objection runs deeper than the former because it focuses on the logical necessity of believing what we take to be true, rather than the contingent matter of not being able to intentionally violate the truth-norm. Of course, the two objections overlap in their criticism of normativism, but they are not the same. Take again the chess example: when we play a game of chess, we necessarily follow the rules of chess, otherwise we are not playing the game. As the normativists say, the rules of chess are constitutive. However, it does not follow from this that when we are in a circumstance to follow the rules of chess, we necessarily *will* follow those rules. It is not, that is, a necessary consequent of me being a position to play chess that *I will* play chess (or even make a single move). Instead, I might

just get up and walk away from the game. In this way, we can take guidance from the rules of chess, even though they are constitutive, and decide just not to follow that guidance. But belief is not like that. Once you are in a position to satisfy the truth-norm, because you think a proposition is true, necessarily you form the belief—you cannot, so to speak, ‘walk away’ from forming the belief; in the relevant circumstances, you must form it. Unlike in the case of ordinary constitutive norms, therefore, we cannot seek guidance from the truth-norm and decide not to form beliefs, in the same way as we can decide not to play a game. Rather, we must form beliefs, and we necessarily do that when we take a proposition to be true.

This is how the no-guidance objection differs from the objection from doxastic involuntarism. The logical relation between us being in a position to form a belief, and us being in a position to obey some other constitutive norm (such as in a game) is different. Once we recognise a proposition is true, we believe it; but once we recognise we can make a move in a game, we don’t necessarily make it—even that sort of minimal guidance from the truth-norm is not permitted when we believe. Moreover, this reveals that the normativists’ appeal to constitutive norms to save normativism does not succeed against the no-guidance objection. Even though the nature of constitutive norms explains why we necessarily must obey norms in certain circumstances (such as when we want to play a particular game), it does not explain why as believers we must form beliefs in the relevant circumstances.

At this point, I consider and reply to Engel’s attempt to avoid the no-guidance objection by interpreting the truth-norm as a normative *ideal*.

5.1.1. Guidance by Ideal

Engel (2013a, c) acknowledges that the no-guidance objection carries a great deal of weight against normative theories of belief. In particular, he goes so far as to concede that

the problem is insurmountable for a *prescriptive* reading of the truth-norm. This leads Engel to take a significant diversion away from how others have interpreted the truth-norm, by making it into an *idealization*. That is, he thinks the truth-norm should be understood as an *ideal principle* (2013a, p. 48) or an *ideal of reason* (2013c). On this understanding, the truth-norm is normative in the sense that it ‘tells us what right belief requires, [but] it does not prescribe that we ought to believe what is true and only what is true’ (Engel 2013c, p. 204). By interpreting the normativity of belief in this way, Engel hopes to have the resources to avoid the no-guidance objection, because idealisations do not provide guidance; at least, that is, not in the way that we would expect from prescriptions. Idealisations are, Engel (2013c, pp. 208-9) tells us, of the *ought-to-be* category of normativity rather than the *ought-to-do* category.

So, the argument goes, because idealisations do not provide guidance as prescriptions do, we should not be surprised that the truth-norm does not provide guidance in that way. However, we should not be too quick to accept Engel’s solution to the no-guidance objection, because when we consider the nature of ideal norms more carefully, we see that the truth-norm fails to provide guidance even in the way typical of ideals.

When developing his position, Engel approvingly cites Hilary Kornblith’s (2001) account of normative idealisations as a way of thinking about the truth-norm as an ideal. Now if we look to Kornblith’s work, we see that idealisations are not without limitations. There are two extremes that idealisations *cannot* occupy. On the one hand, while they need not provide guidance in the same way as prescriptions, they must still ‘in some sense be responsive to human capacities’ such that they can ‘play some role in guiding action’ (p. 238). If this were not the case, and an ideal norm were not responsive to human capacities, it would ‘thereby lose its capacity to play a constructive action-guiding role’ (p. 238). To see what this means, consider again the reading-norm: that you ought to read twelve

novels every day. Now assume that this is an ideal number of books to read, such that *ideally* you ought to read twelve books a day. Does this make the norm more palatable *as* a norm? No, because it is an ideal that we absolutely cannot live up to. So, even when interpreted as an ideal, we must be able to meet up to the standard set by the truth-norm. On this much, Engel (2013c) agrees: ‘I agree with the critics of normativism that the norm of truth has to make a difference. Ideal ought cannot be completely alien to can’ (p. 211). On the other hand, idealisations ‘cannot be so closely tied to what particular individuals are capable of that we fail to recognize that some individuals at some times are incapable of performing in ideal ways’ (Kornblith 2001, p. 238). At this extreme, we see one sense in which idealisations differ from prescriptions. For prescriptive norms, if we say that everyone ought to perform some action, then every *individual* must be able to perform that action. This follows from the *ought implies can* principle as it applies to prescriptions. But for idealisations, *ought implies can* applies in a weaker sense: when considering idealisations, we do not need to ask whether every individual to which the norm applies can satisfy the norm, but only whether the norm can possibly be satisfied in a broader sense, such that someone can satisfy it some of the time.

So, for example, we can make sense of talk about what an individual ought to do or how she ought to behave if she is to be an ideal teacher. Perhaps she ought to have excellent communication skills, be able to give clear and concise explanations of difficult topics, and be willing to put in additional hours to help her students. But that does not mean that every individual is capable of being an ideal teacher. The fact that not everyone can be an ideal teacher does not undermine the fact that ideal teachers ought to act in such and such a way.⁴⁵ For these reasons, normative idealisations provide a target of their own: they are tied to our abilities in a broad sense, such that they must be satisfiable ‘in principle’ (Engel

⁴⁵ Kornblith (2001, §1 & §2) discusses this example, with a particular focus on Feldman’s (2001) interest in the same example construed as a *role ought*.

2013c, p. 209)—but they are not so strictly tied to the abilities of each individual that everyone to which they apply must be able to satisfy them.

This element of normative idealisations allows Engel, first of all, to provide his own solution to the objection from doxastic involuntarism. Interpreted as an ideal, we don't need to be able to choose not to believe the truth, whether we are capable of this is irrelevant. All that matters is that some people are in principle capable of forming true beliefs, and capable of aspiring to the ideal of having true beliefs. Analogously, an ideal teacher need not choose to be an ideal teacher, but may just happen to perform the actions of an ideal teacher which others aspire to—and this is the case even though not everyone is capable of being an ideal teacher. Furthermore, even if no one in fact succeeds at having true beliefs at any particular time, we can still recognise the ideal and aspire to it. According to Engel (2013c), it is enough for the truth-norm to be interpreted as an ideal that 'only certain imaginary beings endowed with powers which are distinct from ours could follow ([for instance] logical saints [and] perfectly rational agents)' (p. 209).

But providing a new way around the objection from doxastic involuntarism is just one advantage of Engel's notion of the truth-norm. What we want to know, however, is whether his account also avoids the no-guidance objection.

The worry for Engel is that, even as an idealisation, there should be some logical space between realising you are in a situation to satisfy the norm and actually going ahead and attempting to satisfy it. This is because, as we have seen, idealisations must still provide some guidance, even if not everyone they guide can satisfy them. But this does not appear to be the case for an idealised truth-norm.

Consider again the example of ideal teachers. Imagine that you want to be an ideal teacher and you decide that ideal teachers speak clearly in class. For this reason, when you are in class, you think about what the ideal-teacher norm suggests, and you try to act accordingly.

In this case, by speaking clearly. While you may fail to satisfy the norm, and may even be incapable of doing so, you can still look to the norm for guidance and opt in or out of trying to satisfy it. This is because there is an inferential step between recognising you are in a position to be an ideal teacher (e.g. by being in class), and acting on the requirements of the ideal-teacher norm—attempting to satisfy the ideal-teacher norm does not follow as a matter of logical necessity of being in a circumstance to satisfy the norm. As with the chess case we discussed earlier you might, for example, just walk away—you may decide in the relevant circumstances just not to teach (or at least not to try and be an ideal teacher). However, for the truth-norm, even if we interpret it as an idealisation, we still attempt to satisfy it as a matter of logical necessity. When we recognise we are in a position such that we take a proposition to be true, necessarily we believe it. So, prescription or idealisation, we are still left without any sense of how the truth-norm guides.

However, with similar considerations in mind, Engel further pushes the idea that the truth-norm *can* guide, but in ways that are not even shared by other ideals (p. 209). To this end, he sets out to demonstrate how the truth-norm influences our belief formation, such that we can still consider it genuinely normative. Specifically, he focuses on certain doxastic practices that we necessarily take part in, and he suggests that that we should think of these practices as ways through which the truth-norm *regulates* our beliefs rather than *guides* them (Engel 2013a, §5).

These doxastic practices Engel (2013c) compares to *decrees of application* that supplement laws in legal systems (p. 212). In his 2013a paper, Engel proposes *transparency* as one decree through which the truth-norm regulates; and in his 2013c paper, he introduces two more: application of the truth-norm in belief formation in favour of rival norms, and our commitment to justifying our beliefs with evidential reasons. I now discuss whether these

three ‘decrees’ are plausible indicators of whether our beliefs are regulated by the truth-norm.

First, consider transparency. As we saw in Section 1.2, transparency is a thesis introduced by Shah (2003), such that the question *whether to believe that p* is answered by, and only by, the question whether *p* is true. Engel’s thought is that our belief forming processes exhibit transparency *because* we are committed to the truth-norm as the ideal standard of reason. Transparency is one of the ways through which the truth-norm regulates our beliefs for truth; and in particular, they are so regulated despite the fact that we do not look to the truth-norm for guidance. This is because, in order for the truth-norm to regulate (via transparency) ‘one does not need to consider the belief antecedently’ (Engel 2013c, p. 212). In other words, for the truth-norm to regulate our beliefs, there is no need for an inferential step between *recognising the truth of a proposition* and *believing that proposition*. The normativity of belief operates in a way that is more direct than other prescriptions or idealisations.

Given this thought, whether we accept Engel’s (and for that matter Shah’s [2003]) proposal, comes down to whether we see transparency in normative terms. Is transparency really the result of our commitment to the truth-norm, or can it be given a descriptive explanation? Those who take the necessity of forming beliefs upon recognising the truth of a proposition as a sign that a normative account of belief is problematic, are unlikely think to think that bringing transparency into the equation makes any difference—surely this is just another observation that tells *against* beliefs being normative. Those like Engel, however, who take transparency to indicate the presence of a truth-norm, are not going to see the necessity of forming beliefs upon discovering the truth of a proposition as a hindrance to a normative account of belief. So, whose thoughts about transparency are really more favourable?

An independent way to decide the matter is to look at cases of transparency in other domains. And in doing so, we find that transparency does not usually indicate a normative presence. The cases I have in mind are analytic truths. For example, consider the stock example *bachelor*. The question *whether S is a bachelor* is transparent to the question *whether S is a married man*, in the sense that answering one of those questions answers the other. However, we don't assume because of this transparency that *bachelor* is a normative concept such that bachelors ought not to be married; and we don't say that bachelors are committed to a non-marriage-norm. Instead, this kind of transparency is just a descriptive fact about bachelors and unmarried men: there is no need to introduce a norm to explain this.

The second way Engel suggests that the truth-norm regulates belief is in relation to rival norms. When face with non-evidential reasons for belief, such as receiving financial rewards to form beliefs, we necessarily recognise the dominance of the truth-norm above these pragmatic reasons. This can be put in terms of the distinction between *right kind* and *wrong kind* reasons for belief. The right kinds of reasons for forming beliefs are those that bear on the truth of a proposition; while the wrong kind of reasons are said to be those that don't bear on the truth of a proposition, namely pragmatic reasons. So, the point is that we cannot, in full consciousness, form beliefs for the wrong kind of reasons; and, according to Engel, this is because the truth-norm necessarily regulates our belief formation, and only right kind reasons for belief are relevant to satisfying the norm. In this vein, he writes: 'The epistemic correctness condition of belief [i.e. the truth] is always the default mode of our belief formation' (Engel 2013c, p. 207).

Again, however, it's not clear why our acceptance of evidential reasons for belief as the right kind of reasons needs a normative explanation. If, as we have seen, transparency does not require a normative explanation, then the fact that only evidential reasons are

right kind reasons shouldn't require a normative explanation either. Instead, the distinction can be considered a straightforward consequence of transparency as a descriptive feature of doxastic deliberation. If transparency is an essential fact about the way we reason concerning beliefs, then inevitably we will only take into account evidential reasons as reasons for forming beliefs. To take other (i.e. pragmatic) reasons into account would not provide us with an answer about what to believe. By analogy, if someone asks whether *S* is a bachelor, and you answer that he isn't, because he has two sisters, then you are supporting your answer with the wrong kind of reason; *S*'s sisters are irrelevant to whether he is a bachelor. But even though there are *right kind* reasons for answering the question whether someone is a bachelor, this still doesn't mean that *bachelor* is a normative concept.

Finally, Engel (2013c) suggests that we again bear witness to the regulation of the truth-norm in our practice of *justifying* our beliefs. He writes: 'The commitment to the truth norm for belief does not entail that [you have] a reason to believe, or a justification. But it entails that [you] should be prepared to give one' (p. 212). For this reason, we should expect and be prepared to give justification for our beliefs due to our commitment to the truth-norm—specifically, in the form of evidence for our beliefs. However, we still may wonder whether this really indicates that we are committed to a truth-norm, given that our practice of providing justification for our beliefs has other explanations. One such explanation concerns the guaranteed success of actions that are motivated (in part) by true beliefs. For instance, if you believe that there is a beer in the fridge, and that belief is true, then when you go to the fridge for a beer, you will get one. But if the belief is false, you are not guaranteed to get one. In this way, having true beliefs is instrumental in satisfying desires, such that having all true beliefs concerning some particular end you might be pursuing guarantees that you will satisfy that end. This is part of the *motivational role thesis* that I defend and develop in Part III, Section 7.1.; but for now, we just need to

see that this fact gives us reason to provide justifications for our beliefs. Why? Because we want to know that when we act on a belief our actions will be successful, and justifying our beliefs with evidential reasons increases the chances of us having true beliefs, and therefore of having successful actions (that is, when those beliefs do indeed factor into action). As such, the practice of justifying our beliefs may have arisen because of our knowledge (if often only implicit) that true beliefs are more useful to us for satisfying our desires than false beliefs. If true, this would mean that we do not need a normative explanation of why we justify our beliefs, but instead just a descriptive explanation of the value we put on having true beliefs.

After reflecting on all the ways Engel suggests that the truth-norm regulates beliefs, we see that none of them are really in need of a normative explanation. Of course, this does not mean that they do not have a normative explanation, but the question I ask is why we should accept there is a truth-norm given all the differences it exhibits from usual prescriptions and idealisations, and given that descriptive explanations can be offered of the ways Engel suggests that belief is regulated by the norm. For these reasons, I hold, the burden of proof is still on the normativist to show how the truth-norm in any plausible sense guides belief formation.

5.2. The Regress of Rules Problem

The second problem that Glüer and Wikforss (2009) raise is the regress of rules objection. The idea behind this objection is that a necessary condition for a rule to be genuinely normative, it must be able to *motivate* us to act. There is a difference, that is, between merely acting *according* to a rule and acting *in accordance with* (i.e. being motivated by) that rule. But, if we accept this condition, then the truth-norm is not a genuine norm, because of a regress that ensues when we apply the norm. Glüer and Wikforss bring the problem to light by focusing on *practical reasoning*. They observe that when ‘*S* is motivated by a rule

R in forming a belief B it is widely accepted that ‘ S ’s forming of B can be at least partially explained in terms of the role R plays in S ’s practical reasoning’ (p. 55). So, if R motivates S to form B a ‘reasons-explanation’ must be available that explains B in part in terms of R . In other words, R must at least partly constitute the explanation of why S formed B . This is simply a consequence of the initial assumption that norms must be able to motivate our behaviour.

Glüer and Wikforss continue: ‘Whatever exactly our model of practical reasoning for rule-following is, in order to be motivated by R , S needs to have a pro-attitude towards what is in accordance with R ’⁴⁶ (p. 55). As such, the initial assumption about norms leaves Glüer and Wikforss with a minimal model of what it takes to be motivated by a rule (or norm) in practical inference:

(P1) You want to act in accordance with rule R .

(P2) To φ is in accordance with R .

(C) You want to φ .⁴⁷

On this model, the pro-attitude is *to want* to act in accordance with R , and φ can be replaced by any action that has the potential to be explained at least in part due to your pro-attitude towards acting in accordance with R . For example, imagine that you want to act in accordance with the rules of a new diet. You learn that drinking kale smoothies is required by the new diet. For that reason, you want to drink kale smoothies. In this case, the diet (i.e. the set of rules that you want to follow) explains, in part, why you want to drink kale smoothies.

⁴⁶ The term ‘pro-attitude’ is taken from Donald Davidson (1963) and refers to attitudes such as desires, wantings, urgings, principles etc., that jointly motivate actions with beliefs. I briefly discuss pro-attitudes in relation to my own theory of belief in Part III, Section 7.1.

⁴⁷ This model is taken from Glüer and Wikforss (2009, p. 55). It helps, I think, to first view the model in the abstract, before applying it to belief. However, in their own discussion Glüer and Wikforss focus in directly on belief and the truth-norm.

Now, notice what happens when R is the truth-norm and φ is replaced with a belief in a particular proposition. Premise 2 becomes: ‘To believe that p is in accordance with the truth-norm’. This creates a problematic inference because ‘such an inference necessarily involves *another belief*, in this case the belief *that to believe that p is in accordance with R* ’ (Glüer and Wikforss 2009, p. 56). Because of this, also the further belief, that *to believe that p is in accordance with R* , needs to be motivated by another rule. But then it follows that the further belief is yet another belief, and the rule for forming this belief is again the truth-norm, so we need to have the *belief that the belief that the belief that p is in accordance with R* , and so on, *ad infinitum*.

This is different from any other cases of rule following we witness, such as in the case of following a diet. In such a case, the reasoning process is grounded in a belief, which itself is sufficient for motivating an action that is in accordance with a rule: a further level of higher-order beliefs, for ordinary rule based practical inference, is not required—while for rules governing belief, such as the truth-norm, there is no foundation. As such, the truth-norm has no motivational force; and in fact, neither does any rule governing belief, as Glüer and Wikforss (2009) make clear: ‘The regress of motivations...poses a fundamental problem for *the very idea of general rules for reasoning*, be they epistemic rules, or whatever’ because it ‘depends on nothing more than the idea that rule-governed performances can be explained in terms of the subject’s attitudes in combination with the idea that belief formation is rule-governed’ (p. 56). Therefore, normative theories of belief are in trouble if they cannot show how the truth-norm (or any other norm) motivates belief formation.

With this second objection from normative impotence outlined, I now show why Engel’s attempt to solve the problem fails.

5.2.1. Engel's Reply

To solve the regress of rules problem from a normative perspective, Engel draws our attention to the *original* notion of doxastic transparency. Unlike the concept *transparency* we discussed earlier, introduced by Shah (2003)—namely, that the question *whether to believe that p* is transparent to the question *whether p is true*—the original version was introduced to answer the question: *How can we know our own beliefs?*⁴⁸ The thought was that we cannot know what our own beliefs are, because to do so we would need to have higher-order beliefs about those beliefs, and yet more higher-order beliefs about those beliefs, and so on—resulting in a regress. However, once we realise that *what we believe* is transparent to (or answered by and only by) *what is true*, we can put a stop to the regress. In order to know what we believe, we need only ask ourselves what is true. As Engel (2013a) puts it:

If belief is regulated through transparency then it is not the case that, in order to answer the deliberative question whether P, one must *believe that one has a belief that P*, in the second order-mode. (p. 57)

In terms of the regress of rules objection, Engel's idea is that—just as we don't need beliefs about beliefs to know what we believe—we don't need to have beliefs about whether our beliefs are formed in accordance with the truth-norm, since we need only ask ourselves what we take to be true to know whether we are committing to the norm:

No second-order belief to the effect that one satisfies the condition expressed by the norm is needed, and it is not necessary to self-ascribe oneself the belief that *P*... in order to be able to answer the question whether *P* (Engel 2013a, p. 57).

But does transparency really help with the regress of rules objection? I think not. The regress involved in rule following does not arise because of the (potential) problem of

⁴⁸ See Evans (1982) and Moran (1988; 2001, pp. 60-5).

being able to self-ascribe beliefs, but because of the problem of determining whether a particular belief follows a rule (in this case, the truth-norm). So, why should knowing what our beliefs are matter to whether we believe that they are in accordance with the truth-norm? This does not make any difference. Even if I know what I believe, by reflecting on what I take to be true, I still require a separate (additional) belief about whether my belief is in accordance with the truth-norm. But if this is the case, then I must also have another belief about whether my initial belief is in accordance with the truth-norm. That is, I must answer the question: *Is my belief that p is true in accordance with the truth-norm?* Yet this leads, again, to the need for more higher-order beliefs about whether our beliefs are following the norm. For this reason, the regress of rules objection still applies: we still need beliefs about whether our beliefs are in accordance with the truth-norm, even if we can know what our beliefs are by reflecting on what we take to be true. Thus, I cannot see how bringing the original version of transparency into the debate helps Engel to avoid the regress of rules objection.

From Section 5, we can see that the two objections from normative impotence hold a deal of weight against a normative account of belief. In particular, we have seen that Engel's attempts to avoid the objections fail. In the first instance—regarding the no-guidance objection—Engel answers by construing the truth-norm as an idealisation—specifically as an ideal of reason. However, despite his attempts to show how the norm regulates belief, such that we can consider it genuinely normative, we are left wondering why we can't just give a descriptive account of the doxastic practices and phenomena he describes. In the second instance—regarding the regress of rules objection—Engel appeals to the original notion of transparency to avoid the problem. Nonetheless, it's not clear how knowing what our own beliefs are helps us to stop the regress involved in knowing whether our beliefs are in accordance with the truth-norm. At this point, therefore, the objections from normative impotence stand.

6. Summary of ‘Norms’

In Part II, I outlined the key elements of normative theories of belief, and discussed some of the crucial objections raised against normativism. The central thesis uniting normativists is that beliefs are governed by a truth-norm, such that we cannot properly account for belief as a distinct attitude unless we recognise this fact. However, as we have seen, this way of characterising belief gives rise to a couple of important difficulties.

Collectively, I refer to these objections, originally raised by Glüer and Wikforss (2009, 2013), as the objections from normative impotence. The first of these problems—the no-guidance objection—draws our attention to the fact that the truth-norm is not like other norms, in the sense that it does not provide us with any guidance about what to believe. This is problematic because we should expect norms to provide guidance if they are genuinely normative; nevertheless, whether we interpret the norm as a prescription or, as Engel suggests, as an idealisation, this worry still applies. The truth-norm just doesn’t seem to have the characteristics of ordinary norms, no matter how we interpret it, which gives us reason to doubt its normative status. Second, the regress of rules problem teaches us that we cannot be motivated by the truth-norm, because to know whether we are acting in accordance with the norm, we must have higher-order beliefs about whether we are, which in turn must be in accordance with the truth-norm, and so on; thus, stripping the truth-norm of any motivational force. Furthermore, Engel’s attempt to avoid this problem by appealing to transparency doesn’t help, because while we can know what our beliefs are without having higher-order beliefs about those beliefs, we cannot know whether our beliefs follow rules without having higher-order beliefs about whether our beliefs follow those rules.

Therefore, I conclude that normative theories of belief, as they stand, fail to give a plausible account of belief. The concept *belief* should not be interpreted as an essentially

normative concept, which in turn means that we cannot answer Hume's Problem by proposing that beliefs are essentially normative. In addition, if we cannot distinguish beliefs from other forms of acceptance on account of their normative status, then we also cannot account for doxastic correctness as a normative notion.

I now turn my attention to outlining and defending my own theory of belief as a uniquely functional device.

PART III
FUNCTIONS

Part III: Functions

7. The Doxastic Effects Theory of Belief

From this point onwards, I develop and defend my own theory of belief. I argue for a functional account of belief, that addresses the three central questions of this thesis. As a reminder:

- (i) What are beliefs? (Hume's Problem).
- (ii) Why do we have them?
- (iii) How should we interpret doxastic correctness?

One of the most fundamental differences between my theory and those discussed in Parts I and II, is that it answers these questions by focusing on the *effects* (or *outputs*) of beliefs, rather than on their *causes* (or *inputs*). As we have seen, teleological and normative theories of belief define *belief* according to the considerations that enter into belief formation: specifically, whether they are formed for truth-conducive reasons (either descriptively or normatively). In contrast, my functional theory focuses on the unique effects that beliefs cause as an attitude, and proposes that we distinguish beliefs according to those effects. I take this general approach from the *traditional motivational role theory* of belief, which characterises beliefs according to the unique role that they play in motivating action.

In particular, for my theory, I defend and elaborate on the traditional motivational role thesis, and I add a second thesis, which I call the *fundamentality thesis*, which also focuses on the effects of beliefs, but this time in relation to other attitudes. Together, I argue that these two theses capture necessary features of belief and are jointly sufficient for distinguishing beliefs from other forms of acceptance—hence, they provide an answer to Hume's Problem.

Central to the first thesis—the motivational role thesis—is the idea that beliefs motivate intentional actions. Traditionally, many thought that the motivational role of belief was sufficient for distinguishing beliefs from other acceptances.⁴⁹ But this was brought into question when Velleman (2000) argued that other acceptances, such as imaginings, can also motivate actions—thus causing many theorists to propose alternative theories of belief, such as the teleological and normative theories. As such, after outlining and developing the motivational role thesis in more detail, I consider Velleman’s reasons for rejecting the sufficiency condition of the thesis, and I argue that we should not be too hasty, given Velleman’s observations, to reject the general spirit of the motivational thesis—which is to distinguish beliefs according to their effects. While, I suggest, the *mere* motivational role of belief may not be sufficient for distinguishing beliefs, it is jointly sufficient with the second thesis.

Central to the second thesis I argue for—the fundamentality thesis—is the suggestion that beliefs occupy a privileged position in our mental framework, in that they are fundamental to our other attitudes, and in particular in their relation to other forms of acceptance. The idea behind fundamentality is that without beliefs we could not have other acceptances, because beliefs necessarily *inform* (in a way later specified) the contents of our other acceptances. If we accept these two theses about beliefs, therefore, we can answer Hume’s Problem.

With regard to the second question, why we have beliefs, I offer a more substantial answer than the theories of belief we have discussed so far. The question in this context is why we, as believers, have come to possess a motivational attitude that is also fundamental to our mental framework in the ways characteristic of belief; why, that is, we have come to possess an attitude that meets the conditions for *belief*. To answer this question, I appeal

⁴⁹ For instance, see Armstrong (1973).

to the literature in philosophy of biology. I propose, specifically, that the motivational role and the fundamentality theses should be interpreted as *functional statements*. This, however, raises an important question: what are *functions*? Answering this is central to understanding the theory of belief I want to provide, so at this point I turn my attention to two influential theories of functions: the *etiological theory* and the *systemic theory*. I argue, for various reasons, that we should accept the systemic theory over the etiological theory. This allows me to offer a systemic account of belief, such that the motivational and fundamental roles of belief are interpreted as systemic functions. With this account given, I am then in a position to answer our second question: roughly, we have beliefs because of the contribution they make, by performing their (systemic) functions, to our evolutionary fitness.⁵⁰

Finally, I focus my attention on the third question, concerning doxastic correctness. I argue that we should, in contrast to the teleological and normative theories, accept a thin (and not substantive) reading of doxastic correctness, such that the correctness of true belief (and the incorrectness of false belief) is not an essential feature of belief, but is instead a pervasive though contingent feature of belief. I suggest that the reason we are strongly inclined to accept a substantive reading is due to one of the aspects of the motivational thesis, which indicates that true beliefs are sufficient for the satisfaction of desires (more on this in the following section). This further commits me to the position that there is nothing essentially wrong with (or defective about) false beliefs, which is not all that surprising when we recognise the various benefits that false beliefs can provide: be these pragmatic benefits, evolutionary benefits, or epistemic benefits.

All things considered, then, I argue that focusing on the effects that beliefs have (in relation to the roles that they play) enables us to distinguish beliefs from other acceptances;

⁵⁰ This statement risks sounding trivial, however my hope is that specifying the *ways* in which belief contributes to evolutionary fitness, by performing its specific functions, alleviates this concern.

I argue that interpreting these roles as functions allows us to give an account of why we have beliefs; and I argue that we should accept a thin reading of doxastic correctness, in contrast to the theories previously discussed.

7.1. The Motivational Role Thesis

Frank Ramsey (1931) is often credited with formulating the precursor to motivational role theories of belief. He metaphorically describes beliefs as ‘maps of neighbouring space by which we steer’ (p. 238). At face value, this metaphor captures the intuition that beliefs guide our actions—that is, at least, when we desire an end that is relevant to our beliefs. As it stands, however, the metaphor is vague; and it isn’t until D. M. Armstrong (1973) that it is fleshed out into a more substantial theory of belief.

Armstrong fully commits to defending a version of Ramsey’s metaphor, writing that it represents his own theory of belief ‘in miniature’ (p. 3). Specifically, Armstrong draws our attention to two crucial aspects of the metaphor: that beliefs are *maps* (of reality), and that they are maps *by which we steer* (pp. 3-4). Let’s consider each of these aspects in turn.

According to Armstrong, thinking of beliefs as maps is to think of the ‘totality of a man’s beliefs at a particular time as a single great map’ which embraces ‘all space and all time, past, present and future, together with anything else that the believer takes to exist’, with the believer’s present self ‘as its central reference point’ (p. 3). Of course, despite the extent of this analogy, the map metaphor is not supposed to make us think of beliefs *literally* as maps. We do not spread our beliefs out on the desk in front of us and see where we are going next. Yet, it remains a useful metaphor, insofar as many important similarities between beliefs and maps do hold. As Armstrong continues: sets of beliefs can be incomplete, just as maps can be; they can be sub-parts of themselves, again like maps; they can be continuously updated as new information comes in; and, crucially, they provide us with reference points for basing action on (pp. 3-4).

However, simply characterising beliefs as maps is not alone enough to distinguish beliefs from other forms of acceptance. As Armstrong recognises, similar analogies can be made with the mere *entertaining* of propositions in (non-doxastic) thought. For instance, a work of fiction can be construed as a set of reference points representing a fictional world. These representations can again be thought of as maps, with the same similarities as those shared between beliefs and maps. Hence, more is required of the map metaphor if it is to be understood as a theory of belief, with the potential to solve Hume's Problem.

This is where the second aspect of Ramsey's metaphor is relevant: beliefs are maps *by which we steer*. The thought is that belief-maps are 'action-guiding' (Armstrong 1973, p. 4); while maps formed from merely entertained propositions are not. We do not, for instance, head south to Dorne after reading G. R. R. Martin's series *A Song of Ice and Fire*. Rather, we make plans according to what we believe. In Armstrong's words: 'beliefs are maps of the world in light of which we are prepared to act' (p. 4). Thus, we begin to see the preliminary elements of the motivational role theory of belief come into fruition. On this account, the relation beliefs have to action is part of what distinguishes them from other acceptances (or, in Armstrong's terminology, merely *entertained propositions*). However, while the map metaphor is instructive, we still need to be more precise about what the relation between belief and action amounts to, if we are to hope to turn it into a complete theory of belief. Assuming Armstrong is correct, we need to be more specific about what it means for beliefs to have an *action-guiding* role.

The first thing to notice is that beliefs do not motivate in isolation. You do not go to the shop because you believe they have milk. You go because you also *desire* milk. That is, you act on beliefs in combination with relevant desires (or other pro-attitudes, such as wantings, urgings, etc., see Section 5.2, fn. 46). As we soon see, this is also a central part

of the motivational role thesis: beliefs are analysed as part of a compound of attitudes that together motivate actions.⁵¹

These considerations lead Neil van Leeuwen (2009) to state what he calls the ‘standard characterization of the motivational role of belief’ as follows:

The standard characterization is that 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 (p. 219).

From this outline, we see that beliefs and desires in combination cause and rationalise action. Together, we can think of the cause and rationalisation of action as the motivating properties of belief. So, beliefs causally influence the way we act, and they rationalise our actions in the sense that we explain our actions in terms of what we believe. For instance, your belief (along with your desire) caused you to go to the shop, but you also explain your action by recognising that you went to the shop *because* you believed they had milk.⁵² In addition, Van Leeuwen’s characterisation also introduces us to the idea that true beliefs, when acted on, guarantee the satisfaction of their counterpart desires. And this much seems true, for instance: when you go to the shop for milk, if your belief that there is milk at the shop is true, then you will get milk. Now, of course, in reality things get more complicated than this. You would also need to have true beliefs about where the shop is, how to get there, etc.—but the basic principle remains the same, if your *set* of relevant beliefs is true, then you will satisfy your desires.⁵³

⁵¹ Later in his book, Armstrong (1973, p. 71) acknowledges that desires also have an important role in motivating actions.

⁵² For a classic defence of the motivational properties of belief see Davidson (1963)

⁵³ For an extended defence of the idea that true beliefs guarantee the success of action see the literature on *success semantics*, and in particular (Whyte 1990, 1997). There is no reason for me to take any stand here on the central thesis of success semantics, that *truth* just is the property that guarantees success. However, I agree that that it is at least one of *truth*’s properties.

The second thing to note about this characterisation, however, is that it doesn't quite capture fully the motivational relation between beliefs and desires. The problem is that, on the current understanding, any relevant belief/desire pair will motivate action; but the fact is we don't just act on the basis of any relevant belief/desire pair. Rather, our desires must be *sufficiently strong* to motivate actions.

For example, if the shop is a fifty-mile hike away and you only desire milk a little bit, then you are not going to head out to the shop. And if your desire is weaker than some other overriding desire, then also you will not go to the shop. So, desires must be sufficiently strong when they combine with beliefs to motivate action—where sufficiently strong is understood as meaning *either* (something like) strong enough to motivate you to act *or* strong enough to override conflicting desires that would prevent you from acting.

As such, it helps to factor into the standard motivational role theory of belief the condition that desires must be sufficiently strong to motivate actions. This yields the following motivational thesis, interpreted now as a theory of belief.

Motivational Role Thesis: An acceptance φ is a belief if and only if, jointly with sufficiently strong desires, φ causes and rationalises actions that will satisfy the desires if the content of φ is true.

According to this thesis, beliefs occupy a unique role among other forms of acceptance in motivating action: causing and rationalising action is both a necessary and a sufficient condition for an acceptance to count as a belief. However, although along the right lines, this conversion of the thesis into a theory of belief is obviously flawed—we need to make a few further modifications.

First, as it stands, for an acceptance to be a belief, it must *in fact* motivate an action. But surely many of our beliefs never factor into action, even though they are still beliefs. We may, for instance, learn a historical fact at school, which we come to believe, but that

never combines with any of our desires in such a way as to motivate an action. The problem then is that motivating action, as a necessary condition for an acceptance to be a belief, is wrong. To resolve this issue, however, we simply need to note that belief *can*, unlike other forms of acceptance, motivate action—that is, when combined with appropriate desires.⁵⁴

Second, beliefs don't always motivate alone, alongside desires. Often, beliefs motivate alongside *other beliefs*, for instance when we go to the shop for milk, we have beliefs about what is at the shop, how to get there, how long it will take, etc. All of these beliefs potentially factor into whether we decide to go to the shop. So, also need to thesis to account for the fact that beliefs often motivate alongside other beliefs. To do this, we can point out that an acceptance is a belief if it can motivate as one of a set of a subject's beliefs.

Third, not just any sufficiently strong desire will motivate an action with just any belief. The beliefs and desires have to be relevant to each other. For example, a belief such as that *you have ice-cream in the fridge* is typically not going to motivate any actions alongside a desire *to do some gardening*. Specifying exactly what kind of relation between belief and desire is necessary for the one to count as the other is a difficult problem, but for our purposes we need only note the epistemic point that *from the perspective* of the subject performing an action, the attitudes need to seem relevant to one another. That is, if a subject takes a belief/desire pair to be relevant to each other, then that is sufficient (in the right circumstances) for being a motivating pair, even if in fact the belief does not turn out to be relevant to the desire.

⁵⁴ I agree here with Sullivan-Bissett (2017), who has recently pointed out that the motivational properties of belief are conditional, such that beliefs *would* motivate action in the right circumstances.

These three points about how beliefs motivate allow us to give a more precise and more plausible formulation of the motivational role thesis as a theory of belief.

Motivational Role Thesis*: An acceptance φ is a member of a subject S's set of beliefs if and only if, jointly with relevant and sufficiently strong desires, and potentially other members of S's belief set, φ can cause and rationalise actions that will satisfy the relevant desires if the content of φ , and of any other beliefs contributing to the motivation of the action, is true.

This more sophisticated thesis provides a more plausible theory of belief. Acceptances are distinguished from beliefs according to a number of distinct motivational properties they possess, that they instantiate when the circumstances arise.

7.1.1. The Motivational Role of Imagining (and other Acceptances)

In response to the motivational account of belief, Velleman (2000) influentially argues that other forms of acceptance do, in fact, share belief's motivational properties. As such, the sufficiency condition of the motivational thesis is false: beliefs must be distinguished from other acceptances according to a different property. This is what leads Velleman to develop his early teleological conception of belief (discussed in Part I). However, I want to suggest that Velleman makes the wrong move to resolve his concern: there is no need to appeal to a theory of belief that focuses on beliefs inputs (such as the teleological or normative theories). Rather, we can characterise belief by continuing to focus on belief's outputs.

Before I explain how we can do this, we need to know more about Velleman's reasons for rejecting the sufficiency condition of the motivational thesis. To make his point, he focuses on (propositional) imagining,⁵⁵ with the aim of showing that imaginings can

⁵⁵ For a discussion of propositional imagining and, in particular, how it differs from imagistic imagining, see Section 1.1.

motivate just like beliefs. In particular, he looks at ‘the context of child’s play, in which imagining disposes [a] child to pretend’ (p. 256). His main example is a child pretending to be an elephant. According to Velleman, the child’s imaginings about being an elephant, and not his beliefs, dispose the child to behave as though he is an elephant. So, when he uses his arms as a trunk to drink from an imaginary pail of water, or when he stomps his feet on the ground because his elephant legs are heavier than his human legs, his actions are motivated by his imaginings of being an elephant, and not by his beliefs. Specifically, his imagining ‘disposes him to behave as would be desirable if he were an elephant’ (Velleman 2000, p. 258). The idea is, in his imaginary world, the child behaves as he would in order to satisfy his desires if he actually were an elephant. That is, as if his imaginings were true. Thus, Velleman’s example is supposed to show that imaginings share their motivational properties with beliefs: imaginings, in the right circumstances, motivate actions in ways that would satisfy our desires if the imaginings were true.

However, I don’t think we should not be too hasty to give up on the motivational theorist’s general approach. Perhaps we can still distinguish beliefs according to their effects. We just need to realise an important point that is overlooked by Velleman: even if other acceptances share belief’s motivational properties, they do not motivate *in the same way* as beliefs.

To see this, we can continue to focus on imaginings. Consider the following two possibilities: (i) that beliefs and imaginings share *exactly* the same motivational role (call this the *comprehensive motivational role thesis*), and (ii) that beliefs and imaginings only bear *some* level of motivational similarities (call this the *limited motivational role thesis*).⁵⁶ If the comprehensive motivational role thesis is false, then as Van Leeuwen (2009) informs us, Velleman’s departure from the motivational theory of belief is unnecessary. This is

⁵⁶ I take these theses from Van Leeuwen (2009, p. 233), except Van Leeuwen refers to the latter as the *vanilla motivational role thesis*.

because there would be room for defining beliefs according to their *unique* motivational role. And, as is apparent, the comprehensive motivational role thesis is, in fact, false. Beliefs and imaginings clearly do not motivate actions in exactly the same way. Just consider what the consequences would be if the comprehensive motivational role thesis were true. For instance, suppose the child who imagines he is an elephant behaves exactly as though he believes he is an elephant. Instead of pretending to drink from an imaginary pail with his arm for a trunk, he would be on his way to the nearest lake to suck water up his nose. Obviously, however, this is not how the child behaves; he does not take on his pretence in the same way as he would if he believed it. So, the comprehensive motivational role thesis is false. We do not behave exactly the same when we imagine a proposition as we would if we believed that same proposition, even if imaginings maintain *some* motivational role in action.⁵⁷

Velleman's rejection of any hope of distinguishing beliefs according to their motivational outputs is, therefore, mistaken. However, while we can reject the view that imaginings share exactly the same motivational role as beliefs, we can still accept the limited motivational role thesis—that imaginings and beliefs share some motivational properties. This is a point that Van Leeuwen accepts, and he sets out to precisely characterise the distinction between belief's motivational role and the motivational role of other acceptances (such as imaginings).

To do so, he establishes (what he calls) the *practical ground relation*, which consists of three theses that depict some of the differences between the motivational roles of beliefs and imaginings.⁵⁸ Now, although I think Van Leeuwen's practical ground relation is more or

⁵⁷ For further arguments against the plausibility of the comprehensive motivational role thesis see: Van Leeuwen (2009, pp. 233-4) and Lucy O'Brien (2005).

⁵⁸ Van Leeuwen's practical ground relation:

1. Attitudes of type *X* are available for motivating actions across all practical settings, while attitudes of type *Y* depend on the agent's being in a certain practical setting to be effective in influencing action.
2. Attitudes of type *X* represent the *practical setting* one is in such that one acts on attitudes of type *Y* on account of being in that setting.

less correct, I won't enter into a detailed discussion of it here, as he overlooks a more concise way of stating the difference, which is to say that beliefs are the most *fundamental* attitude amongst our mental states (including our acceptances). Once this thesis is stated, we can see that it allows us, in conjunction with the motivational thesis, to distinguish beliefs from other forms of acceptance.

7.2. The Fundamentality Thesis

The thesis I defend in this section is that beliefs are the most fundamental attitude in our mental framework.

Fundamentality Thesis: An acceptance φ is a belief only if φ occupies a fundamental role in a subject's mental framework.

The proposal is that beliefs necessarily occupy this fundamental role. The reason this is not also a sufficient condition is because beliefs must also possess the motivational properties discussed in the previous chapter, i.e. to cause and rationalise action—and I'll get to how the motivational role and fundamentality theses relate in the following section. The idea of a *mental framework* refers (minimally) to our acceptances and the relations between them. I say 'minimally' because this relation may hold more broadly between beliefs and other attitudes as well, such as our pro-attitudes; but I won't go into that here. At this point, we need to know what is meant by *fundamental*, which requires some discussion.

By saying that beliefs are 'fundamental' I mean that they possess an important property: that they necessarily *inform* other acceptances. But what does this mean? To say that beliefs necessarily inform other acceptances is to say that we cannot have other acceptances

3. Attitudes of type X are the cognitive input into *choosing* to act with attitudes of type Y as the input into practical reasoning, when one does so choose. In this relation, beliefs are said to be attitudes of type X , while acceptances are Y s. For more information see, Van Leeuwen (2009, §2).

unless we also have relevant beliefs whose content is essential to the formation of those other acceptances. In this sense, the *informing relation* puts beliefs in a fundamental position. In particular, I mean that our other acceptances derive the information contained in their contents from the contents of our beliefs. For example, think again about Velleman's child imagining that he is an elephant. He imagines that his arm is a *trunk* and that he is drinking from a *pail of water*, and he proceeds to use his arm to drink from the pail. But *why* does he imagine a trunk and a pail of water, rather than of a straw and a bottle of coke? It's because he has *beliefs* about what it means to be an elephant and about what an elephant would be doing in a certain context—and his beliefs do not inform him that an elephant would be drinking coke through a straw. Specifically to this case, given that the child imagines drinking from a pail and not, say, a stream, his informing beliefs are perhaps ones that he formed after a trip to the zoo. Of course, this is not to say the child has to imagine exactly what his beliefs tell him about the normal behaviour of elephants—he could, if he wanted, imagine being an elephant drinking coke from a straw, but *still* his beliefs would inform his imaginings, in the sense that he would require beliefs about what straws are, what coke is, and *what it would be like* to be an elephant drinking coke from a straw. The point then is not that we can *only* imagine (or have other acceptances about) what we believe, but that we can only have imaginings (or other acceptances) when they are informed by beliefs.

Furthermore, the claim is that beliefs necessarily occupy this relation to other acceptances, because to have acceptances at all we must have beliefs. For instance, if the child imagines that he is an elephant, he *must* have some beliefs about what it means to be an elephant, otherwise he has no way to form the imaginings. And to give another example, if you assume for the sake of argument that the moon is made of blue cheese, you must have beliefs about what the moon is, what blue cheese is, etc. Otherwise you have no resources to put together the assumption. This is not to say in any such cases that we must have

true beliefs about what we accept. The child might think ‘elephant’ refers to *camel*, and so his imagining would be about the wrong thing (just as his belief is); but he is still able to form his imagining only because he has a belief about what elephants are, even though that belief is false. To drive the point home, imagine right now that you are a Chugga-Du-Wugga. Unless you believe something I don’t, you can’t.⁵⁹

So that is the essential point about fundamentality. Necessarily, beliefs inform acceptances, in the sense that we cannot accept anything without having beliefs to inform the contents of those acceptances.⁶⁰ This means we can state the fundamentality thesis more precisely as a thesis about how beliefs inform other acceptances.

Fundamentality Thesis*: An acceptance φ is a belief only if φ informs other acceptances, such that a subject couldn’t have those acceptances unless φ performed that role.

This thesis more precisely characterises what fundamentality is about. However, there is one final point to make. As it stands, this thesis *requires* beliefs to inform other acceptances. But as beliefs don’t *necessarily* motivate action, they don’t *necessarily* inform other acceptances too. It is quite possible to have beliefs that never have role in informing any of our other acceptances. Thus, one final amendment must be made, such that beliefs must only necessarily possess *the potential* to inform other acceptances. This yields the following, finalised version of the thesis.

⁵⁹ In a different context, Boghossian (2003, pp. 41-45) seems to touch on this aspect of the fundamentality relation, when he suggest that we probably would not be able to have desires if we did not have beliefs about what we desire.

⁶⁰ The argument I offer for this conclusion only focuses on one other kind of acceptance (with a brief mention of *assumptions* toward the end). This may be considered a weak inductive argument to the conclusion that other forms of acceptance are necessarily informed by beliefs. However, I think on reflection it is not difficult to see that other acceptances are in fact informed by beliefs. So, my question to those who doubt the fundamentality thesis is: which kind of acceptance *is not* necessarily informed by our beliefs?

Fundamentality Thesis^{**}: An acceptance φ is a belief only if φ can inform other acceptances, such that a subject couldn't have those acceptances unless φ performed that role.

With *fundamentality* appropriately understood, this thesis captures the unique fundamental role that beliefs occupy in relation to other acceptances. In doing so, it allows us to account, alongside the motivational role thesis, for the differences in beliefs and other acceptances, such that beliefs are the attitude that can perform both a motivational role and occupy a fundamental role in our mental framework.

At this point, it's worth pointing out that I am not trying to deny with the fundamentality thesis that acceptances do, as Velleman suggests, share some motivational properties with beliefs. We can still accept that imaginings, for instance, are what motivate the child to act like an elephant. The point rather is that beliefs occupy a fundamental role in acceptances, such that even though acceptances can motivate, beliefs must always, to some extent, have some role in actions based on acceptances as well. For this reason, we can agree with Velleman (2000) that it would be 'depressingly unchildlike' (p. 256) to give an explanation of child's pretense in terms of only beliefs and desires, as long as we acknowledge that we cannot rule beliefs out of a full explanation.⁶¹

With the considerations of the last two sections outlined, therefore, my proposal is that the fundamentality thesis and motivational role thesis together amount to a theory of belief that solves Hume's Problem.

⁶¹ Nichols and Stich (2003) argue that acceptances do not motivate at all, and suggest instead that *conditional beliefs* are responsible for motivating actions. So, the child pretends to be an elephant because he thinks: 'If I were an elephant, I would act like this'. For all I have said, this could be true, in which case the motivational role thesis is, in fact, sufficient for distinguishing beliefs from other acceptances. However, I see no reason to reject the idea that acceptances can be motivational, even if we agree that conditional (and other) beliefs have an essential role in structuring (in the fundamental sense I have been describing) acceptances.

7.3. Solving Hume's Problem

Combining the motivational and fundamentality theses into a theory of belief yields the following:

Doxastic Effects Thesis: An acceptance φ is a member of a subject S's set of beliefs if and only if:

- (a) jointly with relevant and sufficiently strong desires, and potentially other members of S's belief set, φ can cause and rationalise actions that will satisfy the relevant desires if the content of φ , and of any other beliefs contributing to the motivation of the action, is true; and,
- (b) φ can inform other acceptances, such that a subject couldn't have those acceptances unless φ performed that role.

I call this thesis the *Doxastic Effect Thesis* because it characterises beliefs solely in terms of the effects that beliefs have; in particular, on action and on other acceptances. The motivational properties are, as we have seen, to *cause* and *rationalise* actions, and the fundamentality property is to *inform* other acceptances. No doubt beliefs have many other properties, but my claim is that these properties are necessary features of belief that together are sufficient for distinguishing beliefs from other forms of acceptance. Therefore, the Doxastic Effects Thesis, with its components suitably understood in the way that I have described them above, is my solution to Hume's Problem.

As we proceed, I continue to treat the motivational role thesis and the fundamentality thesis separately, as this helps with clarity. My aim now is to answer the further two questions that are of interest in this thesis; namely, why we have attitudes that meet the conditions set by the doxastic effects thesis (i.e. why we have beliefs), and what we should make of doxastic correctness. To answer the first of these, I now turn to the literature in philosophy of biology on *functions*.

8. Etiological vs. Systemic Functions

In order to interpret the motivational role and fundamentality theses as functional statements, we need to know exactly what it means to *make* a functional statement. To this end, I discuss here two prominent theories of functions: the *etiological theory* and the *systemic theory*.

In essence, the etiological theory focuses on the *causal histories* of functional devices, and holds that functions should be attributed on the basis of biological devices having the appropriate causal history; while the systemic theory holds that functions should be attributed on the basis of the *causal role* that a biological device has in a system—thus, the key difference in these theories resides in whether or not they put emphasis on the *histories* of functional devices. In the following subsection, each of these theories are specified in more detail.

Once they have been outlined, I provide a parallel analysis of each. This analysis looks primarily at a broad distinction in the literature concerning the projects that functional theorists take themselves to be involved in. They typically either consider themselves to be providing a *conceptual analysis* of functions or, alternatively, providing a *theoretical definition* of functions. I don't take a side in this debate, instead I argue that from either perspective, we should prefer a systemic theory.

Broadly speaking, as a project in conceptual analysis, there are some long standing and compelling reasons for rejecting the etiological theory of functions. In particular, the etiological theory leads to a number of counterintuitive consequences concerning how we use *function* as a concept in ordinary language. And as a project in providing a theoretical definition, things are a little more complicated, but considerations again lean in favour of the systemic theory. Specifically, I focus on four theoretical virtues that are generally

considered to be important in the literature, and that we should expect a theory of functions to possess:

(v1) Epistemic Access

(v2) Theoretical Scope

(v3) Theoretical Parsimony

(v4) Explanatory Power.⁶²

As we continue, I explain each of these virtues in detail. However, in short, (v1) and (v2) do not require much discussion, as they are straightforwardly shown to fall in favour of the systemic theory; while (v3) and (v4) are more complicated, and thus require longer treatments of their own.

Concerning (v3), there is an interesting question that looms over the discussion in the functions debate, about whether functions are essentially normative—if they are, then our theory of functions must introduce normative entities into the debate, and if they are not then normative entities are an unnecessary addition. Etiologists assume that functions *are* normative, and so they attempt to account for this normativity. While systemic theorists argue that functions are not essentially normative, and so don't include normative entities in their theory. As such, the theory of systemic functions is more parsimonious, as it eliminates normative entities. However, the question is whether we *need* normative entities to make functional statements, because if we do, then it is no good eliminating them. Thus, I argue in the systemic theorists' favour that functions are not essentially normative, and that attempts by the etiologists to account for the normativity of functions are flawed.

⁶² Surely there are other important virtues that we should expect a theory of functions to possess—however, I choose these because they are often the focus in the current literature.

Concerning (v4), the explanatory power of etiological functions is generally considered to be one of its principle advantages. As we shall see, etiological functions are essentially explanatory, in the sense that attributing an etiological function to a device is (in part) to explain why that device exists. This is contrary to attributions of systemic functions, which are only explanatory in the sense that they explain the role a device has in a system. Thus, as an explanatory theory of why a device exists, the etiological theory wins out over the systemic theory. However, while I agree this is one advantage of an etiological theory of functions, I argue that when interpreted correctly the systemic theory can also contribute to explaining why a device exists, even though the systemic theory is not *essentially* explanatory in this way. Therefore, given all of the other problems with the etiological theory of functions, and granted that systemic theories can contribute to explaining why a biological device exists, I conclude that we should accept the systemic theory of functions—this allows us in the following section to provide a systemic analysis of belief (focusing on the motivational and fundamentality theses), that goes some way to answering our second question: why we have beliefs.

8.1. Etiological and Systemic Functions

The Theory of Etiological Functions

The etiological account of functions descends from Larry Wright's (1973, 1976) influential work. Wright's main contribution to the literature is to suggest that functions are 'intrinsically... explanatory' (1973, p. 154). When we ascribe functions to devices we provide, according to Wright, an 'important kind of explanation' (p. 154). For example, asking why hammers are made the way that they are, and asking what the function of hammers is, are the same kind of question: they ask for an *explanation* of why there are hammers. In general, Wright suggests that function statements can be 'called upon to explain how things *got* there', by which he means: 'something like' why an artefact or

organism has a certain functional device, why that device ‘is where it is’, or why that device exists at all (pp. 156-8). This begins to reveal the *etiological* character of Wright’s theory. Functions of devices, whether of artefacts or organisms, are part of a *causal* explanation of why an artefact or organism has that device. For an artefact, the causal chain runs through the intentions of an intelligent designer: we design hammers with a flat, solid head, because of that feature’s capacity to knock nails into hard surfaces (i.e. its functions); and for organisms, the causal chain runs through the processes of natural selection:

We can say that the natural function of something—say, an organ in an organism—is the reason the organ is there by invoking natural selection. If an organ has been naturally differentially selected-for by virtue of something it does, we can say that the reason the organ is there is that it does that something.⁶³ (Wright 1973, p. 159)

Concerning organisms, one of the many possible examples of functional devices is the auricle (outer ear). A function of the outer ear is to direct sound waves into the auditory canal, because this capacity helps an organism to hear. But *why* is this a function of the outer ear? According to the etiologist, it is because of the contribution that this capacity made to the selective success of past organisms with outer ears. In contributing to the selective success of past organisms, the outer ear increases its own chances of being reproduced in future generations, and therefore is in part the cause of its own existence in current organisms. That is the causal nature of etiological functions in biological devices.

In this way, both artefacts and organisms are attributed functions, on Wright’s theory, according to their etiological history. For Wright, this is the unifying characteristic of functions across the board: whether because of intentional design or natural selection,

⁶³ Wright here appeals to the notion of *selection for* an organ (or biological device). As we later see, selection for biological devices is not, strictly speaking, necessary for attributing etiological functions (Buller 1998). However, for now we need only note that etiological functions are attributed to biological devices via an appeal to evolutionary processes.

function statements are a kind of etiological explanation of why devices exist in the way that they do and of why they exist in a population.

Although, while the unifying nature of etiological functions is disputed, notably by Christopher Boorse (1976), the attribution of etiological functions to biological devices has become dominant in philosophy of biology. Theorists such as Ruth Millikan (1984, 1989, 1993), Karen Neander (1991a, b, 1995, 2006, 2007), and Carolyn Price (1995, 2001) all make significant advances in defending the etiological theory; and in light of a recent objection, Ema Sullivan-Bissett (2016) puts forward an important defence of the theory (discussed in Section 10.2). Of course, their accounts differ in significant respects, but each remains committed to the central idea that the etiological histories of biological devices determine the functions of those devices.

At this stage, we need to be clear about exactly what the central thesis of the etiological theory is. However, as David J. Buller (1998) observes, there is a fundamental ambiguity in the etilogists' proposal: they need to be clear about what they mean by the *selective history* of a biological device, for there are two salient interpretations. Depending on how we interpret *selective history* we can derive two theses that the etilogists could be working with. Accordingly, Buller distinguishes between the *strong etiological thesis* and the *weak etiological thesis*.

Strong Etiological Thesis: The etiological function of a device D of an organism O is to do that which devices of D 's type did to contribute to the inclusive fitness of D 's ancestors, and which caused the genotype, of which D is the phenotypic expression, to be selected for by natural selection.⁶⁴

⁶⁴ This is how Neander (1991a, p. 174) formulates her etiological commitments, though she does not distinguish between the *strong* and *weak* versions of the etiological thesis. For continuity, I have adjusted the variables, and I have changed the term 'proper function' (which is often used by the etilogists) for the equivalent 'etiological function'.

The important aspect of the strong formulation is that, for a device to be ascribed a function, the device must have been *selected for* (which is the commitment we also saw Wright make above). This means, on the technical understanding of *selection for*, that during a device's selective history, there must have been variation in that device, and organisms with that device must have had, in virtue of possessing that device, greater selective success than organisms with variations of that device.⁶⁵ In other words, saying that D was selected for implies that D increased O's *differential fitness*.⁶⁶ Hence, on the strong thesis, functions cannot be ascribed to devices when devices of that type were not selected for; that is, when they did not contribute to an organism's differential fitness. And this may occur, for instance, when a device has no variants in its own selective environment. On the *weak* etiological thesis, however, selection for a device is not required:

Weak Etiological Thesis: A current token of a device *D* in an organism *O* has the function of producing an effect of type *F* just in case past tokens of *D* contributed to the fitness of *D*'s ancestors by producing *F*, and thereby causally contributed to the reproduction of *D*s in *O*'s lineage.⁶⁷

By eliminating selection for as a requirement for function ascriptions, the weak thesis attributes functions 'more liberally' than the strong thesis (Buller 1998, p. 512). All that is necessary for making function attributions on the weak thesis is for a device to have historically contributed to an organism's *fitness*. In contrast to differential fitness, fitness here is not measured in terms of competitive success between organisms with devices of variant types, but rather in terms of whether a device contributed to its own proliferation in future generations *regardless of whether there was any competition*. This means that *D*s can still be ascribed functions even if, in their selective environment, they face no competition

⁶⁵ For this technical understanding of selection for, see Elliot Sober (1984).

⁶⁶ To be specific: *D*₁ in environment *E* has greater differential fitness than *D*₂ in *E* if and only if possessing *D*₁ in *E* gives an organism a higher probability of reproducing than possessing *D*₂ in *E*.

⁶⁷ This is how Buller (1998, p. 507) formulates the weak etiological thesis. I have again adjusted the variables for continuity.

from organisms with variants of D (because, say, there are no such organisms to compete with).

Therefore, the strong and the weak thesis can be described according to two shared conditions, and one distinguishing condition. While the strong and weak theses require:

(i) that a device is hereditary (so it can contribute to fitness); and,

(ii) that the effects of a device *in fact* contributed to the fitness of an organism's ancestors,

the strong thesis also requires,

(iii) that the functional effects of a device contributed to the differential fitness of an organism's ancestors (so that the device was *selected for*).

In this way, the strong thesis entails the weak. Because contributing to differential fitness implies contributing to fitness, if strong function attributions are made so are weak functions attributions. That is, if a device satisfies condition (iii) it also satisfies condition (i).

That concludes our introduction to the theory of etiological functions. However, for my purposes moving forward, the important point to note is that, from here on, I talk solely in terms of the weak etiological thesis when referring to the etiological thesis. This is because, insofar as the criticisms I discuss apply to the weak thesis, they also apply to the strong.

The Theory of Systemic Functions

Like the etiological theory of functions, the systemic theory originates in the 1970s. In this case, Robert Cummins (1975) is responsible for the early philosophical work.⁶⁸ Cummins summarises his theory as follows:

To ascribe a function to something is to ascribe a capacity to it which is singled out by its role in an analysis of some capacity of a containing system. When a capacity of a containing system is appropriately explained by analyzing it into a number of other capacities whose programmed exercise yields a manifestation of the analyzed capacity, the analyzing capacities emerge as functions. (p. 765)

The basic idea is that function statements refer to capacities that yield further capacities in a containing system. A crucial difference therefore between this and the etiological approach is that the systemic theory pays no attention to the selective history of a device. Functions are attributed to devices solely on the basis of the contributions those devices make to the capacities of a containing system. This is the essence of the systemic theory, but recently Paul S. Davies (2000, 2001) has stated the theory more explicitly, and has included an important advance on Cummin's initial characterisation. Thus, Davies' I work with Davies' characterisation from here onwards:

Systemic Thesis: Where "*A*" refers to the analysis of system *S* into components, and where "*C*" refers to the systemic capacity we wish to explicate, device *D* has systemic function *F* if and only if:

- (i) *D* is capable of doing *F*,
- (ii) *A* appropriately and adequately accounts for *S*'s capacity to *C* in terms of the organized structural or interactive capacities of components at some lower level of organization,

⁶⁸ Bock and von Wahlert (1965) give an earlier account functions with systemic elements but without focusing explicitly on philosophical concerns.

- (iii) D is among the lower-level components cited in \mathcal{A} that structurally or interactively contribute to the exercise of C ,
- (iv) \mathcal{A} accounts for S 's capacity to C , in part, by appealing to the capacity of D to F ,
- (v) \mathcal{A} specifies the physical mechanisms in S that instantiate the systemic capacities itemized.⁶⁹

The majority of these conditions Davies derives from Cummins (1975), with condition (v) taken from Cummins (1983, p. 31). However, Davies' contribution is to make systemic analysis explicitly about *hierarchical systems*. We can see how hierarchies play a role in conditions (ii) and (iii) where the focus of systemic analysis is on how the capacities of 'lower-level' components contribute to the higher-level capacities of an analysed system.

As they stand, these conditions are highly abstract, so it helps to demonstrate how they attribute functions with an example. I focus on the patella. An anatomist might want to give an analysis of a human leg (the containing system S), with a focus on the leg's role in enabling us to walk (the system's capacity C). She thus breaks S down into individual lower-level components, such as the leg's muscles, tendons, bones; including the patella (the device D we are focusing on), and studies how the capacities of these components contribute to the leg's role in enabling us to walk. Insofar as these components do have capacities that contribute to our ability to walk, those capacities are *functions* according to the systemic analysis. In our case, the anatomist determines that *aiding knee extension* is one of the capacities of the patella that contributes to walking, so that capacity is thus one of the patella's functions F .

⁶⁹ See Davies (2001, p. 89). I have changed some of Davies' variables for continuity. For Cummins own explicit definition see Cummins (1975, p. 762).

To summarise this analysis: Condition (i) is satisfied because the patella is capable of aiding knee extension; (ii) is in part satisfied by specifying the patella's function, but will only be completely satisfied when all of the remaining functions of the leg's components are specified; (iii) applies because the patella is one of the devices cited in the systemic analysis of the leg; (iv) applies because the analysis appeals to the patella's functional role in enabling us to walk; and (v) will be satisfied once all of the physical mechanisms in the leg that enable walking are specified. Once all of these aspects of the analysis are complete, the anatomist has given a full systemic analysis of the leg, and can say that one of the functions of the patella is to aid knee extension.

According to the systemic theory, then, functions are attributed to biological devices on the basis of the role those devices play in a containing system. In contrast to the etiological thesis, no emphasis is put on the causal history of those devices—it is irrelevant to the systemic theorist whether ancestral versions of a device contributed to the fitness of an organism. My aim in Sections 9-11 is to weigh the pros and cons of the etiological and systemic theories against each other, and to show that in light of a number of pressing concerns, the systemic theory should be preferred.

9. The Project of Functional Theorists

In the literature on functions, there is a dispute about exactly what kind of project functional theorists are involved in. Some take themselves to be offering a *conceptual analysis* of functions, such as Wright (1973, 1976) and Neander (1991a, b, 2007); and others insist that they are giving a *theoretical definition* of functions, notably (Millikan 1984, 1989). In the following, I don't take a stand on which project is correct, rather I argue that from either perspective the systemic theory has important advantages over the etiological theory.

9.1. Conceptual Analysis

Ordinary conceptual analysis (as opposed to *specialist* conceptual analysis, discussed below), involves an attempt to ‘describe the criteria of application that members of [a] linguistic community generally have (implicitly or explicitly) in mind when they use [a] term’ (Neander 1991a, p. 170). When this community encompasses all users of a particular term (or concept), we are able to consider, in giving an analysis of that concept, any way that users of that concept would be intuitively willing to apply it. This approach to conceptual analysis leaves any proposed conditions for the correct use of a concept open to a vast range of hypothetical examples against which the credentials of those conditions can be tested. Interpreted in this way, as a project in ordinary conceptual analysis, the etiological theory runs into some well documented counterexamples.

Two of the more pressing kinds of counterexamples, both originating in Boorse (1976), involve historical examples of function attributions, specifically pre-Darwinian cases; and functional devices coming into existence for the first time, either because of gradual change, spontaneous creation, or ‘unparalleled saltation’ (p. 74).

An example of the first problem is William Harvey’s discovery of the function of the heart, which he determined was to circulate blood around the body, over two centuries before the publication of Darwin’s *Origin*. Given that the etiological thesis depends on evolutionary concepts, such as *fitness*, Harvey cannot have had in mind the etiological thesis when he recorded his discovery. Hence, we cannot say, on the etiological account, that Harvey discovered the function of the heart. And the fact that he attributes to the heart the same function as we get according to the etiological thesis is just a happy coincidence. If therefore our analysis of functions is supposed to be broad enough to include pre-Darwinian cases, then the etiological thesis is false.

Examples of the second kind of counterexamples include *first-devices* and *spontaneous creatures*. Regarding first-devices, the issue is that at some point in evolutionary history, any given device must have performed its function for the first time. However, at such a time, according to the etiological thesis, those devices cannot have been performing functions, properly speaking, because they are not the result of ancestral devices that performed the same function. So, they do not have the appropriate etiological history to be attribute etiological functions. For example, at some point in evolutionary history, an organism possessed a photoreceptive cell for the first time, and used that cell's capacity to detect light to its evolutionary advantage. On the etiological account, we cannot say that this photoreceptive cell performs a function, because it does not have the requisite evolutionary history; but the intuition is that the cell does perform a function, despite not having this history. Hence, the etiological thesis again gives us counterintuitive results, this time concerning first-devices.

Similarly, spontaneous creatures do not have the history required to attribute to their devices etiological functions. One example is Donald Davidson's Swampman.⁷⁰ If a creature such as Swampman spontaneously came into existence, then he would not have any evolutionary history, his devices would not have been subject to the process of natural selection, and we would not be able to ascribe to his devices etiological functions. However, given that he is a replica of an ordinary person, presumably we would be willing to say, according to our ordinary notion of functions, that he has functional devices. We would, for instance, be willing to say that the function of his heart is to circulate blood,

⁷⁰ The Swampman case: Imagine you have gone for a walk by the swamp, but unfortunately for you, you are struck by lightning, while at the same time a tree is struck by lightning. Your body is vapourised and the tree, by some cosmic coincidence, is turned into a replica of you, though out of different molecules (Swampman). Everything about Swampman is like you; his appearance, his behaviour, the structure and organisation of his organs, and his cognitions are all as yours would have been, if you were still alive. The only difference is that he does not share your history, for Swampman has no history (see Davidson 1987, pp. 443-4). Davidson's original reason for introducing Swampman concerns problems relating to *meaning*, not function statements. Nonetheless, Swampman has important implications in the present context.

even though his heart is not the result of evolution. Again, then, the etiological thesis clashes with our ordinary idea of functions, when the project we are involved in is interpreted as one of ordinary conceptual analysis.

In contrast, if we reflect on the systemic theory, we see that such counterexamples do not readily apply. In the case of Harvey's functional statements about the heart, it's reasonable to assume that what Harvey did was provide a kind of systemic analysis of the heart and its role in a containing system. He took a system, the circulatory system, and broke it down into parts; arteries, veins, the heart, etc.; and determined through his analysis that the heart contributes to the system by acting as a pump that circulates blood around the body. That is not to say that Harvey took these exact steps in this order, nor need he have, the point is only that his project more readily fits the conditions of a systemic analysis, and does so without violating any of our intuitions about how we ordinarily ascribe functions.

Furthermore, concerning first-devices and spontaneous creatures, the systemic analysis again meets our intuitions. Given that first-devices contribute to the higher-level capacities of an organism, even though it is for the first time, they can be considered functional. For instance, the first photoreceptive cell may have helped an organism to more easily navigate, thus contributing to the navigational abilities of that organism. As such, on the systemic analysis, the photoreceptive cell can be considered a functional device, despite not having any evolutionary history. Likewise, spontaneous creatures such as Swampman can be attributed functions, because their devices contribute to the higher-level capacities of the creatures to which they belong, in just the same way as the devices of the creatures they resemble. In the case of Swampman, his heart circulates blood, enabling him to survive, in just the same way as an ordinary human's heart. So, we get a

more intuitive result for spontaneous creatures according to a systemic analysis than we do with an etiological analysis.

For these reasons, the systemic theory has clear advantages over the etiological theory when the project is interpreted as one in ordinary conceptual analysis. In cases where a biological device has no evolutionary history, the systemic theory supports our ordinary use of the term *function*, such that we are willing to ascribe functions to that device.

However, not all theorists who involve themselves in conceptual analysis have been deterred by these considerations. Rather than abandon conceptual analysis, Neander (1991a) argues that we should focus our analysis on *specialist language*, thus engaging ourselves in *specialist conceptual analysis* and not ordinary conceptual analysis. Neander tells us that the appropriate way to analyse functions is to pay attention specifically to what ‘contemporary biologists’ have in mind when they make function statements, thus limiting our domain of analysis to contemporary biology (p. 171). From this perspective, the etiological thesis is thought to be immune to the above counterexamples, because the ideas of pre-Darwinian’s and ordinary people about functions, are irrelevant to how function statements are made in contemporary biology.

About pre-Darwinian cases, Neander (1991a) writes: ‘Scientific notions are not static’ (p. 176). They shift with changes in background theory. So, we should not concern ourselves with what biologists of the past had in mind when they made function statements, which they certainly did not make due to a commitment to the etiological thesis. Instead, we should be interested in function statements as biologists use them today, because they attribute functions with modern theoretical background assumptions—such as, that is, evolutionary principles. And concerning spontaneous creatures, from the perspective of modern biology, such occurrences do not happen, so they are not relevant to how biologists make function statements. When contemporary biologists make function

statements, they clearly do not have in mind whether what they are saying would readily apply to a creature, such as Swampman, that just appeared seemingly from no-where, with no evolutionary history. Rather, they are interested in, and form their concepts according to, actual scientific cases, so there is no reason for the scientific notion of function to account for such cases. Furthermore, if creatures were to start appearing out of no-where, biologists would either have to change their current notion of functions accordingly, or they would have to create an entirely new notion of functions; either way, the current notion—the one we are analysing now—is not required to account for such cases. Therefore, according to Neander, the early conceptual problems facing the etiologists do not apply when the analysis is restricted to specialist language. Nonetheless, Neander's suggestions is not free of its own problems

First, even if we assume that Neander avoids some of the previous counterexamples, she does not avoid them all. In particular, first-devices still present a problem. Although creatures such as Swampman are not relevant to contemporary biology, examples of first-devices are. At some point in an organism's evolutionary history, it must possess a device that performs a certain function for the first time, and questions about when and how this device occurred, the conditions in which it produced its function, etc. are interesting questions in contemporary biology. So, first-devices cannot be as easily ignored as spontaneous creatures. They present genuine cases that raise interesting questions in for contemporary biologists.

Second, we can't be sure what contemporary biologists really do have in mind when they make function attributions. One way to find out is to do an empirical study, by producing a random sample of biologists and running a survey. However, Neander rules this strategy out by stressing that conceptual analysis 'cannot be done by deed poll' (p. 176). Her reason

for rejecting this method is that people, including biologists, often use rules for applying concepts that they don't explicitly recognise as rules. To this end, she gives an analogy:

Of course I do not claim that modern biologists have natural selection consciously or explicitly in mind when they use the notion of a "proper [etiological] function"... That a grammatical rule might come as a surprise to us is not proof that we do not employ it, and by the same token, no explicit knowledge is required [by biologists who employ etiological functions] either. (Neander 1991a, p. 176)

The basic idea is that, even though biologists might not realise they are committed to the etiological thesis when they make function statements, which is the contemporary rule for attributing functions in biology, they nonetheless are—just as we commit to grammatical rules even when we don't know what those rules are. But Neander's analogy here is a little bit strange. The fact that we, as ordinary language users, can have grammatical rules attributed to us that we are unaware of, does not show that contemporary biologists, the specialists in their field, can have rules of biology attributed to them that they are unaware of. As ordinary language users are not specialist grammarians, but biologists are specialist biologists. So, given Neander's project—to analyse how specialists use function statements—her analogy does not hold. If she wishes to understand what contemporary biologists have in mind when they make attribute functions, she cannot assume that they must have etiological functions in mind, even if they don't realise it. This approach is simply to insist that what we are trying to prove must be true: contemporary biologists must be committed to the etiological thesis.

If we take seriously Neander's suggestion to analyse how contemporary biologists use function statements, despite what she says, perhaps a survey of what biologists have in mind is not such a bad idea. It is better, at least, than trying to suggest that biologists must have the etiological thesis in mind, even if they themselves are not aware of it.

Finally, short of conducting a survey, another way of determining how contemporary biologists use function statements, is to study the literature in contemporary biology. This is the approach taken by Ron Amundson and George V. Lauder (1994); and although they do not outright reject the possibility that etiological functions are used in biology, they argue that this is nearly always not the case. Instead, they suggest that systemic functions, and not etiological functions, are ‘ineliminably involved in ongoing research programs’ (p. 466). For example, in functional and evolutionary morphology, Amundson and Lauder point out that while theorists often talk about the ‘evolution of function’, they do so without any reference to the effects of natural selection. Rather, they are talking about how the interactions of components in a system change through time. Of course, underlying this change through time are the processes of natural selection, but that is not to say that functions are attributable in virtue of natural selection. Moreover, Amundson and Lauder observe that there is an ongoing shift in functional anatomy towards treating functions as ‘conceptually similar to structures’—language which more readily lends itself to a systemic reading of functions (p. 463). Given Amundson and Lauder’s analysis, then, whether contemporary biologists have etiological or systemic functions in mind is not as clear as Neander would have us believe.

All things considered, despite Neander’s shift in focus from ordinary to specialist conceptual analysis, we are still left without any strong reasons to believe that the etiological thesis is a more plausible theory of functions than the systemic thesis. Moreover, granting that the systemic thesis holds out better as a conceptualisation of our ordinary notion of function, and assuming Amundson and Lauder are correct to suggest that systemic functions cannot be eliminated from contemporary biology, we should prefer, as things stand, the systemic theory as a conceptual analysis of *functions*.

9.2. Theoretical Definition

The alternative project theorists involve themselves in is to provide a theoretical definition of *function*. This is Millikan's (1989) suggestion as she turns her back on conceptual analysis, retorting that it is 'a confused program, a philosophical chimera, a squaring of the circle [and so on...]' (p. 290). Millikan's thought is that the discovery of evolution by natural selection amounts to the discovery of etiological functions, or, to use her own term, of *proper functions*. In this tone, she compares the etiological definition of functions to the scientific definition of, for example, water and gold. Just as water is H₂O and gold is atomic number 79, functions are etiological, properly speaking.

This general approach, as with Neander's, is intended to avoid some of the early objections raised against the etiological thesis. It does so by ruling out the use of function statements that are not etiological, because they do not fit the definition of *function* (just as we cannot refer to *lead* as 'gold' because of how *gold* and *lead* are defined). For example, since Harvey was writing about the heart's function pre-Darwin, he cannot have been making function attributions correctly, because the true nature of functions had not been discovered—and while we might be tempted to attribute functions to creatures like Swampman, this is done in error; an error that is equivalent to finding a clear liquid on Twin Earth and calling it water, even though its chemical formula is not H₂O.

On the surface, this construal of the theory of etiological functions is difficult to dispute. What hope is there in trying to argue against a well-established theoretical definition in science? Once theoretical definitions are accepted in the scientific community, such as the definitions of water and gold, nothing short of a fundamental paradigm shift is going to change them. Moreover, the theory underlying the etiological definition of functions—namely, the theory of evolution—is firmly entrenched in scientific understanding for the

foreseeable future (to put it mildly). Nevertheless, there is still a case to be made against the etiologist.

To begin, unlike other well-established scientific definitions, there is scope for asking whether the etiological thesis really is the correct definition of functions. Given our current scientific background knowledge, it would be naïve to ask whether water really is H₂O,⁷¹ but asking whether functions really are etiological is not naïve, even within the scientific community. This is because the etiological thesis is not a well-established scientific definition of functions; which is, in turn, because the case for a theoretical definition of functions is largely underdetermined.

In biology, whether we adopt an etiological or systemic account of functions, and apply these theories respectively to the capacities of biological devices, we end up making almost all the same functional attributions (albeit for different reasons). For instance, suppose we ask what the function of the mammalian heart is. From an etiological perspective, we compile information about the evolutionary history of hearts, and we determine that our best evidence suggests that ancestral hearts contributed to the fitness of mammals due to their capacity to circulate blood, i.e. their function. And from a systemic perspective, we study how the capacities of the heart contribute to the higher-level capacities of a system (such as the circulatory system), and we learn that these higher-level capacities occur in part because of the heart's ability to circulate blood, so we attribute to the heart that function. Therefore, we make the same function attributions to the heart from either point of view. And on the surface, this holds for other biological devices: one of the kidney's functions is to filter blood, presumably this contributes to the fitness of ancestral organisms with kidneys, but also it contributes to higher-level capacities of the urinary

⁷¹ We might still ask from a *philosophical* perspective whether water is really H₂O, but if we are providing a *scientific* definition, given our current scientific background knowledge, asking whether water is really H₂O seems problematic.

system; the patella's function is to aid knee extension, this contributes to fitness, but also to the higher-level capacities of the leg; and so on. In at least these and many other cases, the etiological and systemic analyses ascribe devices the same functions.

These considerations leave us wondering why we should accept, as Millikan suggests, that the etiological thesis is *the* theoretical definition of *function*. The fact that biological devices are the products of natural selection does not entail that we should attribute functions to them on the basis of their evolutionary histories—nor, for that matter, did Darwin claim to have discovered the 'proper' functions of biological devices. Given, then, that the etiological and systemic theories typically make the same function attributions, we should not be too quick to accept either as the theoretical definition of function. We can, however, ask which interpretation provides a more *useful* notion of functions in various respects. This leads us to a discussion of the various *theoretical virtues* that a theory of functions can exhibit—with preference being given to the theory which is, so to speak, the most virtuous. Four virtues that are generally considered important for a theory of functions are, as I mentioned previously, the following:

(v1) Epistemic Access.

(v2) Theoretical Scope.

(v3) Theoretical Parsimony.

(v4) Explanatory Power.

In the remainder of this section, I focus on (v1) and (v2), while saving (v3) and (v4) for more lengthy discussions in the sections to follow. Concerning the first two virtues, I argue that if we adopt the systemic thesis we are better placed, given the information available to us, to make use of function statements; that is, our epistemic access to the relevant information is greater for attributing systemic functions than etiological functions.

And I argue that the systemic thesis has greater theoretical scope than the etiological thesis, which is especially apparent if we reflect again on some of the earlier counterexamples levelled against the etilogists.

Epistemic Access

Regarding our ability to access the information necessary to make function statements, Amundson and Lauder (1994) discuss at length the difficulties with obtaining the requisite information for attributing etiological functions. They argue that in many instances it is extremely difficult, and often practically (although not logically) impossible, to establish the evolutionary histories of devices to which we wish to attribute functions. The problem is that often the devices are ‘ancient’ with selective histories that extend back ‘hundreds of millions of years’ (p. 461). In these cases, the information we need to determine the selection pressures that were acting on these devices is often completely inaccessible, making it impossible to make etiological statements about those devices with any degree of confidence. Furthermore, as Amundson and Lauder again observe, it is extremely difficult to delineate precisely which device specific selection pressures have acted on—a difficulty that is ‘insurmountable when dealing with fossil taxa or ancient structures’ (p. 461). As such, determining what selection pressures acted on what devices is an incredibly difficult task; and accessing the information required is even, in some cases, impossible. If the etiological thesis is the accepted definition of functions, then we simply cannot make function statements in these cases where the relevant information is impossible to access; at least not without making highly speculative claims.

This does not imply that the etiological thesis is false, but it detracts from the utility of the account, which is especially problematic in science if we wish to have theory of functions that we can readily apply. In this respect, the systemic theory is more appropriate. Although it can be difficult to give a complete systemic analysis of a system,

the information for attributing devices functions within a system is still more forthcoming than the evolutionary history of a device. Experiments can be carried out on organisms, and inferences can be made about the causal roles that devices play in the various systems of the organisms. For instance, we can study the leg by performing experiments that alter or adjust the various components of the leg, with the aim of understanding the causal roles these components have in enabling the leg to perform one of its functions, such as walking. Such experiments don't require any information about the evolutionary histories of devices, but only about how those devices interact to produce various effects. That is not to say that biologists are not interested in evolutionary history, it is just that the information required to make systemic function statements is non-historical in this sense. For these reasons, the problems that come with trying to determine the selection pressures that acted on ancestral devices are not inherited by biologists making systemic function statements.

Given the relative ease of obtaining information for making systemic function statements, then, other things being equal, the systemic thesis is preferable to the etiological thesis. It allows us to ascribe functions to devices without the trouble of making difficult and potentially impossible investigations into the evolutionary history of devices.

Theoretical Scope

The issue with theoretical scope concerns the amount of function attributions a theory of functions makes in seemingly the right places. A theory of functions cannot just ascribe functions to any effects that a device has, but the more instances it can account for that we would usually think of as instances of functions the better. That is, assuming that it also covers the important cases in scientific discourse. In this context, we are brought back again to the kinds of counterexamples that faced the etiological thesis construed as a conceptual truth about functions: cases of pre-Darwinian function statements, first-

devices, and spontaneous creatures. The purpose of providing a theoretical definition of functions was in part, for Millikan, to avoid such counterexamples. However, if we have an alternative theory, such as the theory of systemic functions, that accounts for such cases but also accounts for scientific cases of functions, then surely that is a theoretical advantage of that theory.

To put things differently, we can agree with Millikan (and Neander) that these examples are not relevant to contemporary biology, while also accepting that a theoretical definition of functions is to be preferred if it can account for those cases; assuming, of course, that the theory can also account for the biological cases.

Given that the theory of systemic functions does account for such cases (as we saw earlier in this section), and given that that it also applies in contemporary biology (also this section), we can say that the systemic thesis has greater theoretical scope than the etiological thesis—and as such should be preferred from the perspective of this theoretical virtue.

By focusing on (v1) and (v2)—epistemic access and theoretical scope respectively—we can see that there are reasons for thinking that the systemic thesis is the more preferable theoretical definition of function, at least according to these virtues. That leaves (v3) and (v4)—theoretical parsimony and explanatory power. In the following section, I turn my attention to theoretical parsimony.

10. The Normativity of Functions

Deciding which theory is more parsimonious trades to a large extent on whether functions are *essentially normative*, and as such whether our theory of functions needs to include normative entities. On the one hand, the etiologists think that functions are essentially normative, and they consider it an advantage of their theory that it can account for this

normativity. On the other hand, the systemic theorists don't think functions are essentially normative, and for this reason don't think their theory, or any other theory of functions, must account for the normativity of functions. In this respect, the systemic theory is immediately more parsimonious than the etiological theory, because it posits no normative entities. However, as I mentioned earlier, the important question is whether functions really are essentially normative. If they are not, then we don't need a theory of functions that includes normative entities.

Why think functions are normative? The basic intuition is that biological devices have functions that they are *supposed to* perform. When we talk about the heart, for instance, we don't just say that it circulates blood, but that it is *supposed to* circulate blood; and our eyes don't just see, they are *supposed to* see, etc. From this assumption, we also get the idea that devices can *malfunction*. Given that devices have functions they are supposed to perform, when they don't perform those functions, they are not doing what they are supposed to do. This can occur when a device is diseased, damaged, or defective, such that it no longer performs its function (I'll refer to devices that are diseased, damaged, or defective collectively as *broken*). For this reason, according to the etiologists, we need a theory of functions that explains why devices, even when they are malfunctioning, have a function that they are supposed to perform; that is, we need a theory that explains the normativity of functions.

To this end, the etiologists point out that broken members of a *functional type* share their evolutionary histories with functioning members of that functional type. A patella that does not aid knee extension because it is cracked, for instance, is still a member of the *patella* functional type, because it shares its evolutionary history with other patellae. And it is in virtue of being a member of a functional type that a *token* device has the normative properties of that functional type. So even though a cracked patella cannot perform its

function, it is still supposed to, because that is the function that previous ancestral token patellae performed. Hence the etiological thesis is a normative thesis. When we make etiological function attributions, we also make claims about what functional devices are supposed to be doing. In this way, a broken device is—in a genuinely normative sense—malfunctioning.

The etiologists take this normative element of their theory to speak in favour of the etiological thesis. And in particular, they take it to be an advantage over the systemic thesis which does not have the resources to make sense of the normativity of functions. To be specific, condition (i) of the systemic thesis—that a device must *be capable* of performing its function—entails that when devices cease to perform their function, they are no longer members of a functional type, and no longer have the properties of that functional type. So, even if systemic functions were normative, we could not say that a broken device is not doing what it is supposed to, because in virtue of being broken the device would not possess the normative properties inherent in being a member of a functional type. On the contrary, a broken device on the systemic account just becomes *non-functional* in a descriptive sense.

If we assume, therefore, that a theory of functions must account for the normativity of functions, then the systemic thesis is in trouble. The etiological thesis has the normative resources that the systemic thesis lacks. However, there are a couple of points we can make against the etiologists and the supposed normativity of functions that make a descriptive account of functions, and thereby the systemic thesis, more appealing.

The first point I make questions the very intuition that functions are essentially normative. The second assumes that the intuition is true, but points out that despite recent defences to the contrary, the etiological thesis is in no better position to account for the normativity of functions than the systemic thesis.

10.1. Intuiting Natural Norms

The idea that functions are normative reads normativity into nature. But are there really norms in nature? As we have seen, our intuition favours a positive answer: our biological devices seemingly have functions that they are supposed to perform. Yet it is reasonable to ask where exactly this normativity comes from. What is it about nature that gives rise to the normativity of functions? To answer this question, the etiologists read *purpose* into nature; and specifically, into evolutionary processes. For instance, Millikan (1984) implies that nature, through the processes of natural selection, is trying to achieve some end when she says that ‘nature effectively experiments’ (p. 26); Buller (1998) writes that a functional device has ‘the “purpose” of doing X’ (p. 516); and Price (1995) informs us that the essential feature of a function attribution is to state ‘what a device is supposed to do’ (p. 143).

According to the etiologists, biological devices derive their normativity from *natural purpose*. Nature, through evolutionary processes, has a kind of purpose, such that biological devices are produced to serve that purpose. Insofar as they perform functions that go some way towards satisfying nature’s purpose they are *successful*, in the sense that they are doing what they are supposed to do; and if they fail to satisfy that purpose they are *unsuccessful*, as they are not doing what they are supposed to do. As such, the normativity of functions is the result of the normativity of nature.

However, this appeal to natural purpose is a bit dubious, for it merely shifts the problem back a step. Instead of asking where the normativity of functions comes from, we can now ask what grounds the normativity of nature (or natural purpose).

From a historical perspective, especially looking at the pre-Darwinian era, natural purpose was commonly attributed to divine creation. God designed the natural world and everything in it, and he did so with the intention for biological devices to perform certain

functions, and so that is what they are supposed to do—perform their functions. However, theorists today are generally not satisfied with appeals to divine creation or intelligent design, and instead hold strong commitments to naturalistic explanations. The question then is how do we get purpose into nature without involving a divine creator? At this point, the etiologists appeal to *natural design* (as opposed to intelligent design). The supporting thought, captured here by Philip Kitcher (1993), is that natural selection is a kind of replacement for God: ‘one of Darwin’s important discoveries is that we can think of design without a designer’ (p. 380). Nonetheless, while talking about design is a helpful metaphor in the context, it remains curious why we would take the metaphor in any sense literally, given that Darwin provided the tools for describing nature *without* a designer. Perhaps Darwin’s central insight was to explain how highly complex organisms emerge and diversify from significantly less complex organisms, without the guiding hand of a designer. So, as Darwin did with the designer, why not do away with design altogether?

One reason we might be inclined to hang onto the notion of natural design is our intuition that the complex structures we find in nature *must be* in some sense designed. As Davies (2001) discusses in his various places throughout his book, we are inclined to infer from ‘highly regular and highly complex hierarchical systems’, that those systems are in some sense designed (p. 5). However, it’s likely that this inclination towards natural design is nothing more than the same inclination that historically made us appeal to intelligent design. Yet we were willing to discard creation by a divine designer when we realised it was not necessary to account for the diversity we find in nature, so we should be willing to banish design altogether. To do so is the logical end of accepting that there isn’t a designer—indeed *design without a designer* is simply a confused notion.⁷²

⁷² I agree here with Davies general sentiment:

There is order, regularity, degrees of complexity, and degrees of adaptedness—but no design. Of course, we analyze natural systems into systemic capacities in order to understand and control

Another reason for accepting that nature exhibits design is that it supports the etiologists claim, and our basic intuition, that functions are normative. Accepting that nature does have a natural purpose serves to ground the normativity of functions, which in itself might be thought of as a positive reason to commit to natural design. Nonetheless, the problem with this suggestion is clear. Given that the normativity of functions is what is up for dispute, appealing to it to justify the notion that is supposed to ground it—that is, natural design—is circular. If there are any reasons for accepting natural design, then in this context they must be independent of the normativity of functions.

With these considerations in mind, the intuition that functions are essentially normative looks problematic. I can't see any reason to accept, despite the initial force of the intuition itself, that functions are normative. But where does that leave the intuition itself? It is true, at least, that in ordinary discourse we talk in terms of devices having functions that they are supposed to perform. Yet there is no reason to think that such talk commits us to normative entities. Davies (2001) provides us with a plausible, non-normative, explanation of our intuition. As biological organisms, we are heavily invested in ensuring that our biological devices perform as we want them to. Most of us, most of the time, want to continue living; and to do so, we need our hearts to keep circulating blood, our lungs to keep absorbing oxygen, our kidneys to keep removing waste, and so on. And we also want to experience a certain quality of life, which again requires our devices to keep functioning effectively. Indeed, to desire our devices to keep functioning is itself

their operations, and it is tempting to conceive of these systems in terms of some sort of design. But we are wrong to give in to such temptation. Natural systems are comprised of parts that interact with one another, and sometimes these interactions are astonishingly complex and elegant. But that shows only that natural systems work—they exercise higher-level capacities by virtue of organized lower-level capacities—without having been designed and without exemplifying marks of design. Darwin did not show us how to understand the world in terms of design despite the absence of a designer; he showed us instead that we ought to stop thinking of the world in terms of design. (Davies 2001, p. 14)

conducive to, and the result of, natural selection. Such desires are going to incline us towards taking actions that we at least believe will ensure that our devices continue to function. We typically are not, that is, going to perform actions that we know are detrimental to the functioning of our devices. As such, our desire to keep our devices functioning is so strong and ingrained into us that it is easy to make the psychological leap from *understanding how our devices must function to keep us alive* to the idea that *our devices are supposed to function in some particular way*. Nonetheless, the fact we are inclined to make this leap does not mean that function attributions are normative.

Therefore, arguing as the etiologists do, that a theory of functions must be able to account for the normativity of functions, based on our intuition that functional devices are supposed to perform their functions, is incorrect. There are no grounds for reading normativity into nature, because there are no grounds for assuming that nature has a purpose nor that the natural selection is a kind of designer. Furthermore, the intuition itself has a plausible explanation in our inclination to strongly desire our biological devices to continue to function. As such, there is no problem with accepting a theory of functions that is essentially non-normative, such as the theory of systemic functions.

Putting these concerns aside for now, next I question the credibility of the etiologist's claim that the etiological thesis can, in fact, account for the normativity of functions. In particular, I focus on a recent defence of the etiologist's account of normativity by Sullivan-Bissett (2016), which she raises in response to a pressing objection.

10.2. Etiological Functions and Normativity

One of the professed advantages of the etiological thesis is, as we have seen, that it can account for the intuition that functions are normative. Contrary to this claim, however, Davies (2000, 2001) argues that the etiological thesis in fact fails to account for the normativity of functions. This is because, according to Davies, when a device is broken,

it doesn't have (or no longer has) the normative properties associated with functional members of an (etiological) functional type. As such, we cannot say that broken devices are malfunctioning according to the etiological thesis, because there are no grounds for claiming that they are supposed to be doing what members of a functional type do. In more specific terms, Davies' argument can be reconstructed as follows:

(P1) A token device *D* in an organism *O* is a member of an (etiological) functional type if and only if:

- (i) *D* in *O* has the correct evolutionary history (as per the etiological thesis); and,
- (ii) *D* possesses the functional property that contributed to the inclusive fitness of *O*'s ancestors.

(P2) A token *D* in *O* fails to acquire or loses the property that contributed to the inclusive fitness of *O*'s ancestors if *D* is broken.

(P3) Tokens of *D* in *O* that are broken are not members of a functional type. (From P1.ii and P2.)

(P4) Tokens of *D* in *O* possess a normative property only if they are members of a functional type.

(P5) Tokens of *D* in *O* that are broken do not possess a normative property. (From P3 and P4).

(P6) Tokens of *D* in *O* (normatively) malfunction only if they possess a normative property.

(C) Thus, tokens of *D* in *O* that are broken are not malfunctioning. (From P5 and P6.)⁷³

⁷³ I derive this argument from Davies (2000), but also see Davies (2001, Ch.7-7.3). In his discussion, Davies focuses on both the strong and weak etiological theses (on this distinction see above, Section 8.1). However, for continuity, I immediately restrict my construal of his argument to the weak etiological thesis. This should not be a problem, given that the argument can easily be modified to focus on the strong etiological thesis. The only alteration required is to add also to P1.ii that the device must have been *selected for*.

With the exception of (P1.ii), the etiologists should have no problem committing to any of the above premises.⁷⁴ Being a member of a functional type is what grants a device the normative properties associated with that functional type. For example, members of the *lung* functional type have the function to *absorb oxygen*, so token lungs are supposed to absorb oxygen, assuming that functions are normative. However, if Davies' argument is sound, then he shows that, in fact, the etiological thesis does not account for malfunctions, because broken devices are not members of a functional type. Thus, one of the most compelling reasons for accepting the etiological thesis—that is accounts for (normative) malfunctions—is false.

Sullivan-Bissett (2016) accepts the weight of Davies' objection, but she denies that the etiological account cannot be redeemed, such that it can account for malfunctions. She proposes that when the etiological thesis is appropriately extended it can account for malfunctions, while remaining true to its central etiological commitments. In particular, she argues that for a device to be a member of a functional type, the device does not need the capacity to perform its function, but instead requires a set of *intrinsic structural properties* specific to ancestral tokens of that device. The thought is that natural selection does not only target fitness enhancing properties, but also the 'underlying heritable physical configuration responsible for possession of the survival-enhancing [property]' (p. 9). According to Sullivan-Bissett, this physical configuration is the 'primarily heritable' feature of a device, meaning that the survival enhancing property is 'only heritable derivatively, by virtue of the heritability of the physical feature that gives rise to [it]' (p. 9). Interpreted this way, physical configurations (or intrinsic structural properties) are centrally important to the evolutionary history of functional devices. As such, they are

⁷⁴ Davies (2000, §II) argues that the etiologists must commit to (P1.ii), even though they don't specifically endorse that condition. I won't go into the details of Davies' argument for this premise here, as I am more interested in Sullivan-Bissett's response, and she takes no issue with (P1.ii).

part of what it means to be a member of a functional type; and as Sullivan-Bissett writes: ‘possession of the physical [a] physical feature need not necessitate possession of [a] survival-enhancing one’ (p. 9). Essentially, the idea is that, structural properties can grant token devices membership to a functional type, even if that device no longer performs its function. More precisely, Sullivan-Bissett’s suggestion is as follows:

Etio-Structural Thesis

A token device D is a member of a functional type T and has F -ing as its (etiological) function if and only if it:

- (a) Possesses one of a set S of intrinsic structural properties, $\{p_1, p_2, p_3, \dots\}$
- (b) Stands in a causal historical (etiological) relationship to at least one token device 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^* of properties responsible for that token [device] F -ing, $\{q_1, q_2, q_3, \dots\}$
 - iii. contributed to the inclusive fitness of ancestral organisms by F -ing.⁷⁵

On this extended etiological thesis, to be a member of a functional type, a token device must have the appropriate evolutionary history (as per the original etiological thesis); but no requirement is put on a current token device to still be able to perform the function of its ancestors. Instead, membership of a functional type is granted to current token devices in virtue of their intrinsic structural properties. For example, a token heart is a

⁷⁵ See Sullivan-Bissett (2016, p. 10). For continuity, I have changed some of the variables. I have also altered condition (b.iii) to reflect the weak etiological thesis. Sullivan-Bissett’s original condition reads ‘iii. was selected for F -ing’, which demonstrates a commitment to the strong etiological thesis. However, as I discuss in Section 8.1 above, the strong thesis entails the weak, so there should be no problems with discussing Sullivan-Bissett’s proposal in terms of the weak etiological thesis.

member of the functional type *circulates blood*, because ancestral hearts circulated blood, which contributed to the inclusive fitness of ancestral organisms, and because it shares with ancestral hearts a set of structural properties.

The unique thing about the etio-structural thesis, then, is its appeal to intrinsic structural properties as determining factors for whether a device is a member of a functional type. But does the addition of conditions (a) and (b.i) to the original etiological thesis really help the etiologists attribute normativity? I think not.

A crucial problem with this proposal is that the structural properties referred to in the etio-structural thesis collapse into further functional properties. However, in virtue of being further functional properties, they cannot be appealed to in order to avoid Davies objection, because, as we know, functional properties are at the heart of that objection. To demonstrate, I begin in the abstract, and show why the structural properties are really just functional properties in disguise, before considering Sullivan-Bissett's more specific proposal—that the relevant structural properties might be interpreted as the processes of *gene expression*—and showing again why this is just amounts to a further appeal to functional properties.

According to the etio-structural thesis, any 'one of a set' of structural properties shared between a current token device and an ancestral device of the same type is sufficient for the device to be a member of a functional type. If, let's say, the relevant set of structural properties for human lungs is $\{x, y\}$, then any token lung that possesses either property x or property y , provided that the lung also has the correct evolutionary history, is counted as a member of a functional type associated with lungs; such as, the *absorbs oxygen* type. However, imagine that the presences of x or y alone is not sufficient for producing anything that resembles a lung, but something more akin to what Millikan (1984) calls a 'glob of misplaced organic matter' (p. 25). Then only properties x and y when they are

together are sufficient for doing the structural work required for producing lungs. If this is the case, then we don't want to say that a glob of misplaced organic matter is a member of a lung functional type; it may be nothing at all like a lung. For something to be considered a lung, whether malfunctioning or not, it must more closely resemble a lung. So, just any one of a set of shared structural properties is not sufficient for granting a device membership of a functional type. If, like in our case, one property acting alone produces an unidentifiable glob of matter, then we cannot say that the device is in the same category as it would be if it had additional structural properties. In the lung case, for instance, one property is not enough to make a lung. Rather both properties x and y are required, otherwise we don't have anything like a lung.

This means an adjustment must be made to the etio-structural thesis, one that is more specific about the core set of structural properties required to make a device sufficiently like other devices with which it shares some of its structural properties. Thus, for instance, in the lung case, the required set of structural properties to make a lung must include both x and y ; only then do we have a device sufficiently lung-like. But this means the etio-structural thesis must commit to the stronger condition that the relevant structural properties must include those properties which together are sufficient for structuring a device such that we can recognise it as a device of a certain type.

Although this point is not reflected in the etio-structural thesis, later in her paper Sullivan-Bissett does seem to have something like this stronger condition in mind when she says that the required structural properties must be 'within a range considered a reasonable attempt to produce [a device] similar to ancestral [devices] which performed F' (p. 10), and where this range captures 'those [devices] which come pretty close to being well functioning members of a given kind, but nevertheless go awry' (p. 11). These additional specifications enable Sullivan-Bissett to rule out mere organic globs as members of

functional types, while including devices that are not performing their function. However, as I have said, this is a stronger commitment than appears in conditions (a) and (b.i) of the etio-structural thesis—which in turn leads us to the main problem, which is that these structural properties now collapse into functional properties of their own kind.

The problem is that once we start talking about structural properties that are within a ‘reasonable range’ of the properties required to produce a particular device, we are talking about properties that have been *selected for* producing that device (assuming that they produce a device that itself was selected for)—and when a device is selected for producing a certain effect, that effect has necessarily increased the fitness of an organism, and is therefore a *function* on the etiological thesis. In the (hypothetical) lung case, the lung possesses the set of structural properties $\{x, y\}$, precisely because those properties were selected for producing lungs.

To some extent, Sullivan-Bissett acknowledges this point; she writes that the structural properties referred to in the etio-structural thesis delineate a kind and that ‘[p]ast instances of this kind were selected for in virtue of their role in the production of [devices] which were able to perform the function of F-ing, which is adaptive’ (p. 11). However, she doesn’t seem to realise that structural properties which were selected for producing devices meet the conditions for being functional properties on the etiological thesis. Just like any other function, these structural properties have effects (in this case producing devices) that contribute to the inclusive fitness of organisms that possess those properties. So, the etiological chain, so to speak, runs through Sullivan-Bissett’s structural properties, turning them into functional properties of their own kind.

But we can’t rely on further functional properties to grant membership of a device to a different functional type, because these further functional properties would in turn be subject to Davies’ objection. Rather, the only kind of properties that are going to

successfully do the work that Sullivan-Bissett wishes to do are non-functional properties, which her structural properties unfortunately are not.

Now, if we ask more specifically when Sullivan-Bissett thinks a device's intrinsic structural properties are, we see that the problem still applies. Concerning what these properties are, Sullivan-Bissett remains decidedly neutral, suggesting that 'there are no doubt numerous ways tied in with the production of functional [devices] which might delineate the set of [structural] properties' (p. 10). However, she does pick a likely candidate for these properties in the processes of gene expression.⁷⁶ In particular, she proposes that we might appeal to 'an appropriate range of gene expression associated with recent ancestral [devices] which performed F' (p. 11). On this interpretation, devices that have been produced through a sufficiently similar process of gene expression are members of the same functional type (assuming also that they had the correct evolutionary history). But we need to remember that this capacity for gene expression is just as much subject to the processes of natural selection as the capacities of our other devices. As such, we can give an etiological account of the devices responsible for gene expression (e.g. genes and regulators), with their function being to produce proteins which in turn constitute further functional devices. Presumably, such an account would be more or less correct concerning the evolutionary history of gene expression, because this capacity is integral to increasing the fitness of ancestral organisms, given the role that capacity plays in ensuring that we have organs.

In this way, gene expression is, like other functions, responsible for its own existence in a population, in the etiological sense. Genes and regulators cause the process of gene expression to take place, which in turn causes us to have functional devices, which in turn contributes to the inclusive fitness of organisms, which in turn means those organisms

⁷⁶ 'Gene expression' refers to the process by which information in a gene, in combination with transcription regulators, is used to produce gene products, such as proteins.

have an increased chance of reproducing themselves. Thus, on the etiological account, that process of gene expression turns out to be a functional type (just like *absorb oxygen*); only it is a function that occurs at the micro and not the macro level.

The problem of structural properties collapsing into functional properties also, therefore, holds when we consider Sullivan-Bissett's more specific proposal. Moreover, if the role of structural properties is, as Sullivan-Bissett says, 'to produce' devices, then it is difficult to see how this problem can be avoided no matter what the structural properties are supposed to be. For these reasons, the etio-structural thesis, in extending the etiological thesis to include structural properties, fails to avoid Davies objection: we still have no reason to grant broken devices membership of a functional type, and as such no way to account for (normative) malfunctions.

Reflecting again on the parsimony of the etiological and systemic theories, the etilogists give us no reason to accept either (a) that we need to include normative entities in our theory, or (b) that the etiological thesis is capable of accounting for such normative entities. This means that the systemic theory, from the perspective of parsimony, should be preferred to the etiological theory, as it does not appeal to unnecessary (and problematic) normative entities in order to attribute functions. Once again, then, the theory of systemic functions wins out regarding theoretical virtues. That leaves us with one final virtue to consider.

11. The Explanatory Power of Functions

This final virtue—explanatory power—is important for my thesis; not only as a deciding element between the etiological and systemic theses, but because the explanatory power of function attributions enables us to answer the second central question of this thesis: why we have beliefs. The basic idea is that, if function attributions help us to explain why a device exists (recall Wright), then interpreting the doxastic effects thesis as a functional

statement is going to help us explain why we have beliefs. So, we must ask, in what sense are the etiological and systemic theses explanatory?

The etiological thesis, in this respect, is clearly explanatory. When we attribute an etiological function to a device, we necessarily make a claim about the causal history of that device. For example, when say that a function of the eye is vision, we are claiming that vision is one of the reasons that current organisms have eyes; and this is so because vision contributed to the fitness of ancestral organisms with eyes, thus increasing the chances of descendent organisms having eyes. If we accept the etiological thesis, then, we have a theory of function statements that is essentially explanatory. This is not something I am going to dispute here. I agree that *if* we accept the etiological thesis, then we have an explanatory theory of functions. However, I have disputed many other elements of the etiological thesis above, so for those reasons I don't think we should accept it. Instead, my main focus in this section is on elucidating how the systemic thesis can also be explanatory—although I agree that, unlike the etiological thesis, it is not *essentially* explanatory.

In contrast to the systemic thesis, the explanatory power of systemic functions is not typically thought to be in their ability to explain why devices exist. Rather, they explain the causal roles that devices play in higher-level containing systems. For example, the heart's (lower-level) function to circulate blood (in part) causally explains the circulatory system's (higher-level) function to distribute oxygen and nutrients around the body. The circulatory system is able to perform its function because of the heart's ability to perform *its* function. In this sense, the systemic thesis is explanatory, but in the sense of explaining why devices exist at all, the systemic thesis is not explanatory—by attributing a systemic function to a device we make no claims about why that device came to exist as a device.

However, for my purposes, I suggest that at least some systemic analyses of functional systems go some way to explaining why particular devices exist in those systems.

To demonstrate, we first must recall a crucial aspect of the systemic thesis. As we saw earlier (section 8.1), Davies introduces to the notion of systemic functions the idea that systemic analysis only applies to systems with hierarchical structures. The basic components of a systemic analysis, in this sense, require us to analyse higher-level functions in terms of lower-level functions (or vice versa). This is crucial because, as Davies (2001, p. 6) observes, systems we may want to analyse—like human organisms—need not be individuals, but may also be populations. When the system is a population the immediate lower-level components are, thus, the individuals in that population. As such, one of the capacities of a population we may wish to analyse is how that population remains stable in a specific environment, where stability is understood as that population maintaining reproductive efficacy in that environment. In these terms, a complete systemic analysis of a system, where the system is a stable population, can be used to give an explanation of why individuals in that population have the functional devices that they do. That is, insofar as those devices are understood to make a causal contribution, up the causal hierarchy, to the stability of the population.

To clarify by means of an illustration, imagine that the system we are analysing is a stable human population in an environment E, and the device we wish to explain the presence of is the heart. The analysis, broken down into relevant components, is depicted in Figure 1 below.

In the diagram, the square boxes represent systems or devices; the numbers represent functional types, in particular those relevant to the capacities of the system immediately above—for instance, 5 may represent the type *circulates blood*, while 1 may be *social behaviour*; and the triangles represent tokens of their associated functional types, as it is the tokens

that cause the effects that contribute to the stability of the population up the hierarchy. As we can see, the analysis in Figure 1 is only partial, and a full systemic analysis of a human population will include many more systems, devices, and functions; including those of other individuals in the population, such as x_2 and so on (which I completely omit from Figure 1). But as it currently stands, if we assume that the correct devices and functional types are specified from the *heart box* to the *stable population box*, then the diagram at least provides a systemic analysis of the role the human heart plays in ensuring that a human population remains stable.

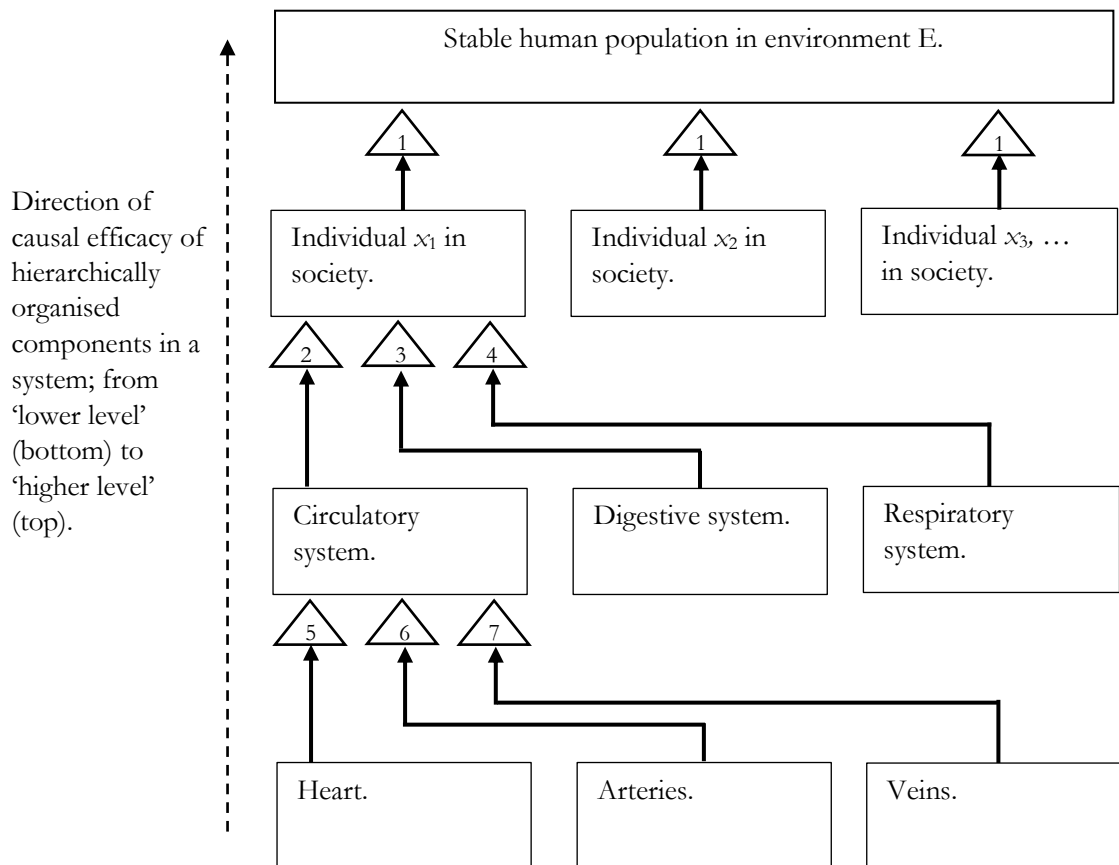


Figure 1
Systemic Functions and the Heart

Making function attributions in this analysis does not require us, as the etiological thesis does, to make any claims about the evolutionary histories of the functional devices. But the rejection of this requirement does not prohibit us from pointing out that the system

we happen to be analysing is one that, in virtue of being a stable human population, reproduces individuals of its own kind, which in turn continue to maintain the stability of the population (all other things being equal). In other words, we are able to specify that the system is *self-perpetuating*.⁷⁷ The significance of this point is great. For, it allows us to say that in virtue of the functions that devices have within a self-perpetuating system, they are in part responsible for their own existence; because they cause the reproduction of further individuals that also possess those devices. To put things differently, by contributing to the functions of higher-level systems, lower-level devices, in producing their functions, are in part responsible for the reproductive capacities of the population. Thus, a *complete* systemic analysis of a self-perpetuating system helps us to explain why functional devices exist in that system—they do so because of the causal contribution they make to the stability of the self-perpetuating system.

To return to the example of the heart. In human individuals, the heart contributes up the hierarchical chain to the functioning of the circulatory system, and the circulatory system in turn contributes to the individual as a whole, such that she can survive and reproduce. When this individual is then in a population that is capable of reproducing members of its own kind, various interactions between the individuals (both biological and social), and between the individuals and the environment, enable the population to remain stable, and thus for the individuals to continue to further reproduce individuals. These further individuals will then also have hearts, which then also contribute to their survival and the stability of the continued stability of the population (other things being equal). As such, we are able to explain why individuals in a human population possess hearts—and an analogous systemic analysis can be inquired into for any of our other biological devices.

⁷⁷ I take this point from Davies (2001, p. 153) who points out that systemic analyses can apply to self-perpetuating systems (although this is not necessarily so).

Finally, I want to make clear that by appealing to the self-perpetuating nature of human populations, I am not making functions essentially etiological in nature. I am not, that is, bringing the etiological thesis in through the backdoor. On this account, function attributions are still made without any necessary appeal to evolutionary histories of functional devices; the point is just that if a system does happen to be self-perpetuating, as human populations are, then we can rely on this self-perpetuating aspect of the system to help us to explain, in unison with a systemic analysis, why devices exist. For instance, even if we have no idea about evolutionary processes, we can still, on this systemic account, correctly attribute to the heart the function to circulate blood, because we can still recognise (as Harvey did) that this is the capacity of the heart that contributes to the (higher-level) functioning of the circulatory system. It's just that *given* that we do understand evolutionary processes, we can also add to the analysis that human populations reproduce in such a way that they remain stable in certain environments *and the systemic function of the heart is (in part) the cause of this stability*.

To summarise this section, attributions of systemic functions to devices in self-perpetuating systems, such as human populations, help us to provide explanations of why those devices exist in those systems. Although this explanatory power is not, as it is with etiological function attributions, an *essential* feature of the theory, we can still see that in the right context, this is a feature of systemic analysis. So, given all the previous considerations against the etiological thesis, and given that the systemic thesis can be explanatory in the relevant sense, we should accept the systemic thesis over the etiological thesis—and this remains true whether the project we are involved in is to provide a conceptual analysis of functions or a theoretical definition of functions. Given these considerations, I now turn my attention to offering a systemic account of the components of the doxastic effects thesis; an analysis which enables us to explain why we have beliefs.

12. A (Systemic) Functional Theory of Belief

My aim in this section is to apply the method of systemic analysis to each component of the doxastic effects thesis: i.e., to the motivational thesis and the fundamentality thesis respectively. The reason for this is to then give a systemic account of why we have attitudes that satisfy the conditions of these theses; that is, an account of why we have beliefs.

Although the two theses are united under the doxastic effects thesis, they remain distinct in terms of the roles they attribute to beliefs—meaning that on my functional account they are distinct functions that beliefs have. Accordingly, I provide two distinct systemic analyses of these functions. We must remember at this point that attributing more than one function to our beliefs is not problematic. Other biological devices can and often do have multiple functions. For instance, the function of the mouth is to chew, but it is also to facilitate speaking.

12.1. A Systemic Analysis of the Motivational Role Thesis

We can begin by recalling the motivational role thesis from Section 7.1:

Motivational Role Thesis*: An acceptance φ is a member of a subject S 's set of beliefs if and only if, jointly with relevant and sufficiently strong desires, and potentially other members of S 's belief set, φ can cause and rationalise actions that will satisfy the relevant desires if the content of φ , and of any other beliefs contributing to the motivation of the action, is true.

As previously discussed, this thesis provides a lot of useful information about the role beliefs plays in relation to desires and action. But before giving a systemic analysis, we need to be more specific about how beliefs motivate in contrast to desires. The way beliefs motivate, to return to Ramsey's metaphor, is by providing us with 'a map of neighbouring

space'. Our beliefs, in this respect, serve to *guide* our actions. We can, so to speak, *look* to our beliefs and determine the *route* that we need to take (at least as it seems to us). *Desire*, in contrast to belief, stands in here as a placeholder for what Davidson (1963) calls a *pro-attitude*. Other pro-attitudes include wantings, urgings, principles, and prejudices. These attitudes also motivate our actions, but in a different sense to beliefs. They don't guide our actions, but they are also responsible for causing them. While a belief alone does not motivate us to act without a relevant pro-attitude, a pro-attitude alone does not cause us to act without a relevant belief. They must motivate together, though they do so in their own ways. We can say, then, that a belief's motivational role is to provide *guidance* while a desire's motivational role is that of a *pro-attitude*. These points help us with the systemic analysis.

A first step in this direction is to say that, taken individually, one of the functions of our belief is to provide guidance, and one of the functions of our desires is to act as a pro-attitude. These functions interact, on the motivational thesis, to motivate action—and action can in turn be thought of as a higher-order function which a belief-desire pair contributes to producing, when they are taken as a system of interacting attitudes. This is analogous to, say, thinking of the muscles in the hand as each having their own function, such that when these functions interact they produce higher-level functions of the hand, such as grasping objects. Interpreted in this way, we can begin to form a systemic diagram, similar to the example analysis of the heart we encountered in Section 11.

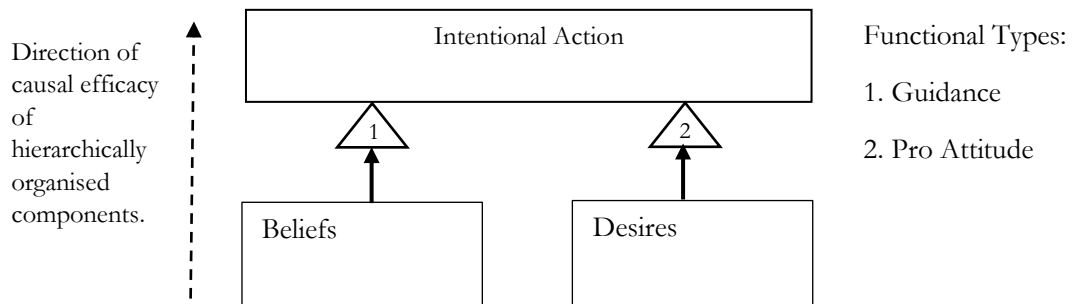


Figure 2
Systemic Functions and the Motivational Thesis
141

In the diagram, we can clearly see how beliefs and desires function together to produce intentional action. Of course, not just any beliefs and desires function together; as we discussed in Section 7.1, they must be relevant to one another. But the point is that being able to perform this function is part of what it means for an attitude to be a belief.

As it stands, this is sufficient for a basic systemic analysis of belief. Beliefs have the lower-level functions of providing guidance, and when this function interacts with that of a pro-attitude such as a desire, this produces the higher-level function of intentional action. However, the motivational thesis allows us to be more fine-grained than we are with this analysis, in that it also specifies that *true beliefs* guarantee the successful satisfaction of relevant desires (when they are acted on). This is important because it allows us to extend our analysis up the hierarchical structure, such that we have an insight into how the function of our beliefs contributes to the stability of our (human) populations.

To see this, we need to recognise, as many authors have, the adaptive benefits of having true beliefs. For instance:

Creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind. (Quine 1969)

I am assuming that, as a general matter, it tends to promote survival and reproduction to have roughly true beliefs about one's environment... If a Mack truck is about to hit you, and you believe that a Mack truck is not about to hit you, this will not on average tend to enhance your reproductive fitness. (Street 2009, 235)

And this is at least partly right. Insofar as beliefs relate directly to action, and assuming that our desires are usually directed towards adaptively beneficial actions,⁷⁸ then we will

⁷⁸ I discuss this assumption about desires in this section, below.

increase our chances of reproduction by acting on true beliefs. And in virtue of increasing our chances of reproduction, true beliefs are contributing to the stability of human populations.⁷⁹ This enables us then to produce a more specific systemic analysis which focuses on true beliefs and how they result in successful action, and thereby the continued stability of a population.

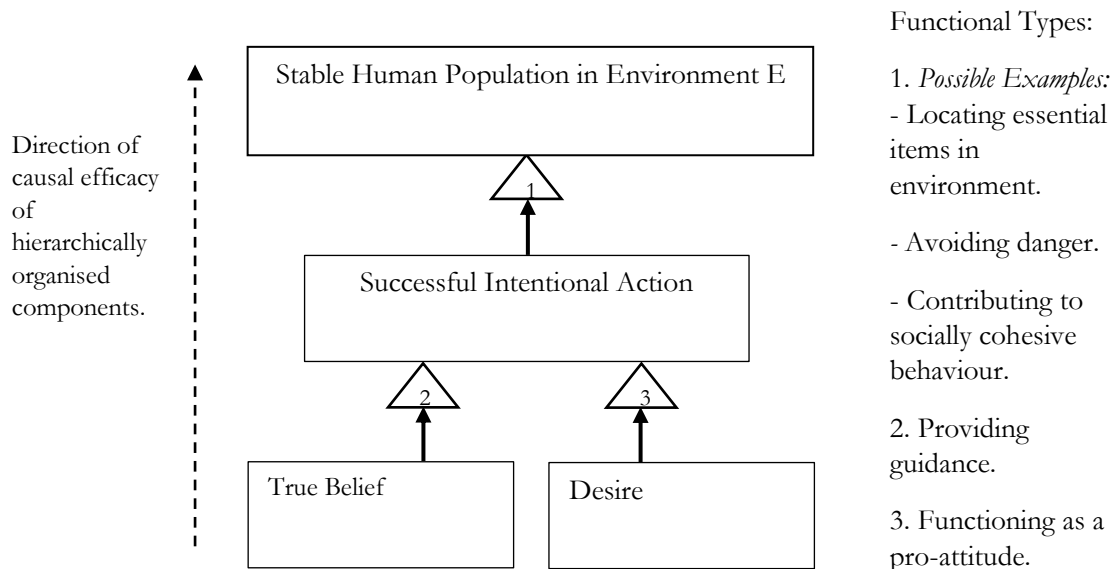


Figure 3
Systemic Functions and True Belief

This representation depicts the function of true beliefs together with desires to produce successful action, and it states that the successful action of individuals in a population results (on the whole) in the stability of that population (where stability, recall, refers to the ability of individuals in that population to reproduce). Now, we need to note that this analysis depends on an important assumption independent of our discussion of belief.

⁷⁹ Many theorists, such as Millikan (1984), take the success of true beliefs in relation to intentional action to be a reason for thinking that all beliefs are regulated for truth. This however is not true. Some beliefs never related to intentional action, and when false can be adaptive for different reasons nonetheless. So, we have no reason to think that all beliefs are regulated for truth. In particular, many evolutionary psychologists think we have adaptive misbeliefs, which are beliefs that have adaptive benefits when false. I briefly discussed adaptive misbeliefs in Section 1.2. I also discuss them again later in Section 13.

This assumption is that our desires—an attitude we have not looked at in any depth—are mostly directed at things that are conducive to evolutionary success. Because of this, the kinds of things we desire are usually ends that increase our chances of survival (as a species). And, although I won't say much about this claim here, it is plausible. For instance, we desire food when we need nutrition; we desire to avoid dangerous situations; we desire to look after our young and protect those close to us; we desire sex; and so on. All of these things play into the kinds of actions we take, and all have an impact (whether in the short or long term) on evolutionary fitness. That is not to say that people never desire destructive ends, such as when an individual desires to commit suicide, but rather it is to say that *most of the people most of the time* desire ends that are in some sense conducive to evolutionary fitness, which helps to support our assumption.⁸⁰

If we accept this assumption, then, true beliefs paired with relevant desires contribute to the stability of human populations. Examples of how actions function to produce this success are listed in Figure 3 under Functional Types, number 1.

Thus, we have a more fine-grained picture of the function of belief in relation to the stability of human populations. True beliefs in combination with desires motivate successful actions (as per the motivational thesis), and successful action causally contributes to the stability of human populations. Later, in Section 12.3, we see how this analysis goes some way to accounting for why we have beliefs at all.

Before continuing, however, there is an important issue that needs addressing that concerns the function of belief in Figure 3. It might be thought that because *true* beliefs

⁸⁰ From an evolutionary perspective, it hasn't historically mattered that some people some of the time have destructive desires that they act on. All that has mattered is that sufficiently many people act on desires that keep the probability of survival high enough for that population to survive (of course, this may change if one individual with destructive desires had a great deal of power and was able to destroy the human race). Also, the fact that desires can be destructive is not an argument against them generally being adaptive—by analogy, the hand can be used to grasp a weapon for killing oneself, but this is not an argument against the adaptive value of being able to *grasp*.

are responsible for contributing to the population's stability, only true beliefs come out on this analysis *as* beliefs. The thought here is that functions of beliefs are necessarily defining features of beliefs. This would be an unacceptable consequence of the above proposal because it would entail that false beliefs are not really beliefs.

The answer is to remember that the systemic analysis in Figure 3 is only a partial analysis of belief that focuses specifically what makes beliefs successful in relation to action. This analysis picks out a property that beliefs sometimes (but not necessarily) possess, that guarantees that those beliefs will satisfy relevant desires, but it alone does not tell us what beliefs are. The motivational properties of belief as described in the motivational thesis (i.e. that they cause and rationalise action), are the defining features of our beliefs (see Section 7), and these properties do not require beliefs to have the success property (i.e. truth). So, beliefs are beliefs in virtue of their motivational properties, but the success property in the motivational thesis allows us to expand our systemic analysis such that we can begin to understand how beliefs contribute to the stability of human populations. By analogy, we can focus a systemic analysis on what makes a heart successful, insofar as it contributes to the stability of a population, thus observing that hearts circulate blood; but this does not mean that non-functioning hearts are not hearts. Therefore, false beliefs are not denied the status of beliefs just because they do not guarantee the successful satisfaction of desires, and in a broader analysis such as in that depicted in Figure 2, we can see that false beliefs can still possess the motivational properties required of beliefs.

12.2. A Systemic Analysis of the Fundamentality Thesis

The basic idea behind fundamentality, as described in Section 7.2, is that beliefs occupy a fundamental role in our mental framework. Like the motivational role of belief, this fundamental role can be interpreted as a systemic function. Recall the thesis,

Fundamentality Thesis**: An acceptance φ is a belief only if φ can inform other acceptances, such that a subject couldn't have those acceptances unless φ performed that role.

This thesis does not focus on the role of belief in relation to action, but the role of belief in relation to other acceptances. The fundamental role of belief is to inform other acceptances, in such a way that we require beliefs to have other acceptances. That is, we must have attitudes with this fundamental property in order to have other acceptances (again, see Section 7.2). But how do we interpret this thesis as a functional statement?

One of the troubles with this approach is that a different systemic analysis can be given for each different form of acceptance, insofar as each acceptance has its own unique role in our mental framework. For instance, assuming that p for the sake of argument has a different role than imagining that p , and imagining that p has a different role again than guessing that p . So, as we work up the systemic hierarchy, each of these difference forms of acceptance bear different relations to the individuals that have them, and moreover the populations to which these individuals belong. This means that for an exhaustive systemic analysis of our mental framework, focusing on how beliefs relate to the stability of human populations up the hierarchy, we would need to specify the unique functions of each form of acceptance other than belief. However, that would be a mammoth task which I cannot hope to complete here.

Another related problem is that a systemic analysis of the fundamentality thesis is going to be speculative insofar as we *don't* know the specific functions of acceptances besides beliefs. Hence, to make progress, we have to assume that at least some other acceptances *do* contribute to the stability of human populations in some way, even though exactly how they do is going to depend on which acceptance we choose to focus on.

With these concerns in mind, I offer then a tentative systemic representation of the fundamentality thesis by focusing on only form of acceptance: (propositional) imagining. This at least is enough to show how we can interpret the fundamental role of belief as another of beliefs functions, and to demonstrate how this role *can* contribute to the stability of human populations up the systemic hierarchy. The systemic diagram is provided in Figure 4, below.

This diagram shows beliefs contributing to the production of an instance of imagining, by informing and regulating that imagining. For example, the child imagines he is an elephant, his beliefs about what elephants are inform the content of his imagining, and they also regulate how far he is willing to take his imagining (for further discussion see Section 7.2).

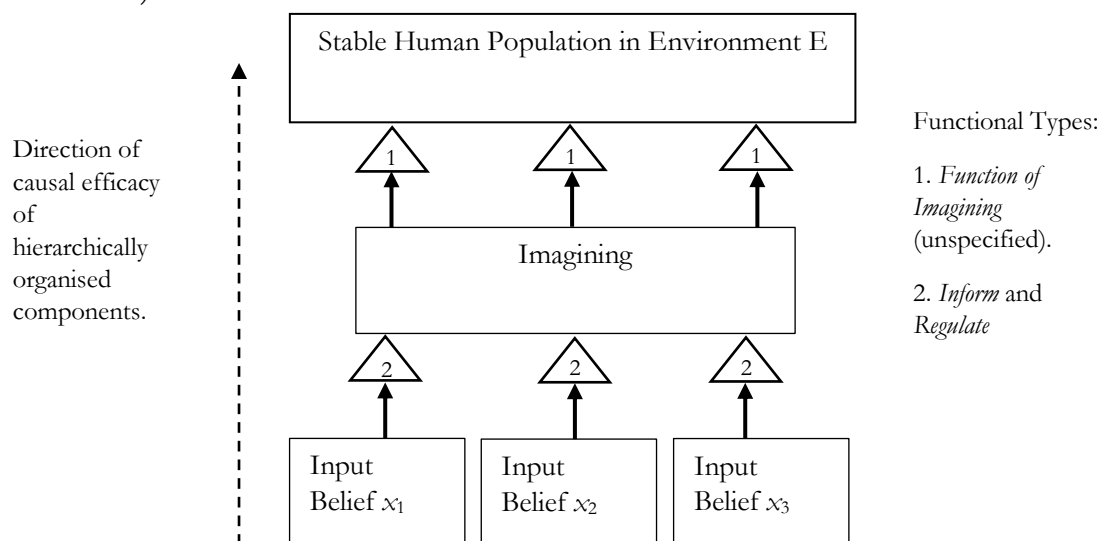


Figure 4
Systemic Functions and the Fundamentality Thesis

As we can see, this diagram depicts three individual beliefs as inputs into an instance of imagining; but there is not necessarily a limit on the number of beliefs that contribute to a given instance of an acceptance. The point of the diagram, however, is to show how beliefs can function together when informing and regulating an imagining—it is not intended to show exactly how many beliefs are required to perform those functions. Going up the causal hierarchy, we then get the speculative (and highly simplified) proposal

that imagining contributes to the stability of human populations (and the same claim could be made for other acceptances). The thought is that our capacity for imagining has arisen as an adaptation, such that imagining, for whatever reason, has increased our chances of survival and reproduction as a species. I won't defend this claim in detail here, but it does at least strike me as plausible. A few reasons for thinking so are the potential roles imagining has in the production of art, in helping us to plan, in our social relations such as when we empathise with others; and it also may aid our mental health, such as when we imagine things to be better than they are, etc. If these examples are suggestive, then we can agree that imagining has adaptive benefits. Thus, on this assumption the causal chain is completed up to the stability of human populations. We begin with beliefs at the lower-level, and the analysis shows how beliefs, because of their function as the fundamental form of acceptance, contribute to the higher-level capacity of a human population to remain stable.

As we see in the following section, making these connections helps us to show why we have beliefs at all. And in particular, to show why the function of belief is in part responsible for the continued presence of beliefs in human populations.

12.3. Why We Have Beliefs

In the above two sections, we have seen how the motivational and fundamentality theses can be interpreted as functional statements about belief. The importance of providing these analyses is that, if they are correct, then they give us an insight into explaining why we have beliefs—and thus help us to answer the second central question of this thesis.

To understand how this explanation is going to work, we can reflect back again on the example systemic analysis of the heart in Section 11. As we saw, by performing their function (to circulate blood) hearts are in part responsible for their own existence in a population, because their function causally contributes up the systemic hierarchy to the

stability of the population. Specifically, this is because the higher-level system (the population) is a self-perpetuating system, meaning that the contribution that hearts make to the system is also a contribution to ensuring that more individuals with hearts will be produced in the population.

Similarly, then, the contribution that belief makes due to its functions to the stability of human populations increases the likelihood that descendent individuals within those populations will continue to possess beliefs, if my analysis is correct. Simply put, in virtue of their capacity to believe, individuals will be more likely to reproduce individuals of their own kind, which too have the capacity to believe.

Looking more specifically at the functions of belief, in the case of the *motivational function* of belief, beliefs along with desires motivate actions. When those beliefs are true, those desires are satisfied, and the ability to satisfy desires is, on the whole, adaptive. These adaptive benefits in turn mean that beliefs contribute to the stability of human populations by causing action, and are therefore partly causally responsible for the reproduction of more human individuals with the capacity to believe. Thus, one of the reasons we have beliefs is because of their motivational function.

In the case of the *fundamental function* of belief, we see that beliefs inform our acceptances, such that they are required for us to have further acceptances. These acceptances are then interpreted as adaptive, which in their own ways contribute to the stability of human populations. Thus, because of their fundamental function in the mental framework, beliefs again causally contribute to the reproduction of further individuals in a population, and therefore to their own continued existence in the population.

Generally speaking, as long as we can create a causal chain from a device, up a systemic hierarchy, to the stability of a human population, we can give an account of why that device continues to exist in the population. And this can be done for belief when we have

delineated beliefs functions, such that we can see their relation either to adaptive behaviour (as in the motivational function) or to other acceptances which themselves have adaptive benefits (as in the fundamentality thesis). As such, we have reached an answer to the second question of the thesis, by interpreting beliefs and their roles as devices and their functions. However, before moving on, there are a couple of potential queries that need clearing up.

First, these explanations of why we have beliefs do not, I admit, explain why we have beliefs *in the first place*. The above systemic analyses why descendent individuals in a population continue to have beliefs, but they do not at all explain why this capacity arose to begin with. While this is true, I do not see it as a serious concern for my systemic account of belief. To find out why we developed the ability to believe in the first place, we would need much more information about the conditions of early human (and probably pre-human) development. And this would not only need to be about the environments of early human populations, but also about the early evolutionary development of the mental capacities of individuals in those populations. Probably, however, we will never be able to access this information, and so will never be able to give a definite account of why our mental capacities appeared in the beginning. Moreover, no other theories that I know of provide an account of why we had beliefs in the first place. Even the etiological theorists, whose theory is inherently explanatory, in the sense that etiological function attributions are said to explain why devices exist, do not profess to show why those devices came into existence. They only explain, as the systemic theory can, why those devices continue to exist in a population. So, the systemic theory goes at least as far as the etiological theory in explaining why we have beliefs, and therefore it should not be seen as a disadvantage of the theory that it does not explain why we humans had beliefs to begin with.

Second, perhaps beliefs also have other functions, besides those discussed above, which contribute to the stability of human populations. For example, adaptive misbeliefs seem to provide adaptive benefits, but not necessarily due to their motivational or fundamental functions.⁸¹ No doubt this is true, and beliefs do perform other adaptive functions. However, my aim here has not been to delineate every adaptive function of belief, but rather to analyse those that are relevant for distinguishing beliefs from other forms of acceptance, such as to answer Hume's Problem (see Section 7).

13. Doxastic Correctness

In Part III of this thesis, I have so far drawn together resources from the literature in order to provide a theory of belief that answers two of the central questions of this thesis: what beliefs are (Hume's Problem), and why we have them. That leaves one final question to address: how to interpret doxastic correctness.

We can clearly state doxastic correctness as follows:

It is correct to believe that p if and only if p .

The important thing to remember when deciding what to make of this principle, is that the functional theory of belief I propose defines beliefs solely according to the effects (or outputs) of beliefs, and makes no essential claims about their causes (or inputs). This is significant, because in contrast to the teleological and normative theories of belief, nothing follows from my theory about the correct or incorrect *reasons* for belief, such that to believe for truth-conducive reasons is the correct thing to do, and to believe otherwise is the incorrect thing to do. Instead, the emphasis is on the functions that beliefs perform, and hence how well they perform those functions.

⁸¹ I briefly discuss adaptive misbeliefs in Section 1.2., they occur again in Section 13.

Using the terminology I stated in the Introduction, this means that doxastic effects thesis does not itself provide a *substantive* reading of doxastic correctness, such that there is an essential teleological or normative relation between belief and truth, but rather a *thin* reading, such that doxastic correctness does not state an essential truth about beliefs. This may seem to be a counterintuitive result at first, but it does not have to be. In the thesis, we are provided with an explanation of why it is that we come to give so much weight to doxastic correctness; that is, we can account for a thin reading of doxastic correctness, such that it has arisen as a pervasive, though socially constructed principle. To do so, we need to focus on the motivational thesis, and in particular the condition that true beliefs guarantee the successful satisfaction of desire when acted on. Following this, I provide some cases of false beliefs that are not in any sense defective, when we understand beliefs according to their effects.

13.1. True Belief as Correct Belief

As I mention in the Introduction, the thought behind the thin reading of doxastic correctness is that ‘true belief’ is used interchangeably with ‘correct belief’, as well as ‘false belief’ and ‘incorrect belief’; however, this is not because of an essential teleological or normative relation between belief and truth, but is a non-essential social construct. So, my question in this section, is why these equivalences have emerged, and why we are inclined to read them in the substantive sense.

The reason, I suggest, is contained in the motivational thesis. Even if only implicitly, most of us possess a kind of commitment to the motivational thesis. We know that if we act on a true belief, we will satisfy our desires; but if we act on false beliefs, something can go wrong—we might fail. For instance, if we want to get food, get places, entertain ourselves, perform well in our careers, etc. having true beliefs about how to satisfy these desires is going to ensure that we do satisfy those desires, but false beliefs can take us in the wrong

direction—and this recognition that true beliefs outperform false beliefs in terms of their relation to action has a strong impact on how we view the relation between belief and truth.

Most of the time we want to satisfy our desires, so we want beliefs that will ensure that we can satisfy our desires. So, we have a strong inclination for favouring true beliefs over false beliefs, because of their utility. Even when true beliefs don't factor into an action, and might never do, we may still be inclined to prefer true beliefs *just in case* they are ever a part of the motivation for an action. For instance, you might have lots of beliefs about certain historical facts which you doubt you will ever need to employ, but you might desire them to be true nonetheless, just in case they are ever required to help satisfy some desire (say, in a pub quiz)—and recognising this even implicitly can have a strong influence of how we view true beliefs. In many, possibly most, cases—and especially the more obvious cases in which beliefs directly motivate action—they are more valuable to us.

In addition, this recognition extends in to the public domain. When we are in discussion with other individuals, and in particular in a debate about what the correct thing to believe is, we know that having true beliefs has important consequences in relation to how people are going to act. We assume, that is, that people want to act on beliefs that will enable them to satisfy their desires—and true beliefs have this role. This is just to say that we want to be appropriately informed when we are trying to determine what the correct course of action is. Thus, in discussion, because of the importance of true belief in relation to action, we accept that as a society we are aiming, on the whole, to have true beliefs. In this way, doxastic correctness arises as a social construct, due to an acceptance that we share, concerning the benefits of having true belief, in relation to action. Conversely, we can see the same observation reflected in the practice of lying. When we lie to someone, often it is because we do not want them to succeed at satisfying some certain end of theirs

that we suspect them to be pursuing (as when we tell the mad axe murderer that our friend isn't home). As such, knowing that a true belief will further the other's end, by giving them the resources to satisfy their ends, we attempt to make them believe something false, because false belief does not come with the same guarantee of success. It is our implicit acceptance of the motivational thesis gives us reason to value true above false beliefs, and in turn is reflected in our social practices.

Therefore, because we acknowledge that true beliefs lead to success, we come to accept doxastic correctness as a principle about belief. And because we value successfully satisfying our desires so highly, our commitment to doxastic correctness strikes us as something substantial about our beliefs. We are inclined to think that true beliefs are always correct, and false beliefs are always defective, for whatever reason. But given that we are concentrating on the effects of belief, and not their inputs, this commitment need not be so. Besides the obvious benefits of having true beliefs in relation to action, there are also benefits to having false beliefs in many situations; and it is likely that false beliefs in some context even outperform true beliefs in the benefits they yield to us—although not benefits that relate directly to the satisfaction of desires. I now consider some of these benefits.

13.2. Non-Defective False Beliefs

Given that true beliefs are valued for their effects (they motivate successful actions), there is no reason, according to the doxastic effects theory, to hold that there is anything essentially defective with false beliefs when their effects are not detrimental to our ends; and, in fact, we may hold that false beliefs are more valuable than true beliefs when they do have beneficial effects.

First, a belief may never perform any of its functions, even though it has the potential to. Imagine having a belief, say, that Alexander the Great enjoyed his tuition under Aristotle.

This belief may plausibly never be required to either motivate any of your actions, or to inform or regulate any of your other acceptances, yet you may hold the belief nonetheless. In this case, given that the belief has no effects at all, it is really quite irrelevant, from the perspective of the doxastic effects thesis, whether the belief's content is true or false. Either way, the belief has no positive or negative consequences. So, if the belief is false, there is really nothing essentially defective about it in terms of its functions—it doesn't perform any. True, it doesn't meet our socially determined standard of correct belief, but besides that we have no grounds for saying that it is essentially defective. It just doesn't make sense to talk about the belief that way. And I am here echoing Papineau's (1999, p. 25) thoughts that there are pragmatic values to believing truth (as per the motivational role thesis), but that when a belief has no pragmatic bearing, there can be no harm done, and thus nothing defective about, a false belief.⁸²

Second, individuals may benefit from having false beliefs when they have higher-order desires that run contrary to their immediate desires. Consider a smoker who has recently quit, and who has a higher-order desire not to smoke, but *in the moment* has a strong desire for a cigarette. The smoker may go looking for his cigarettes in the kitchen cupboard, because he believes he left an old packet in there, only to discover that his belief is false—there are no cigarettes in the cupboard. Here, the smoker's false belief benefits him because it helps him to satisfy his higher-order desire not to smoke. He may, of course, if his immediate desire is strong enough, continue to look for cigarettes elsewhere; but at least for the time being he is unable to smoke because of his false belief.

It may be thought in this case that there is a sense in which the false belief is not beneficial to the smoker, as it does not help him satisfy his immediate desire for a cigarette. In that

⁸² Papineau (2013, §5) does, however, think that there is a 'sense in which' false beliefs are incorrect, but he thinks, in line with what I call the thin reading of doxastic correctness, that this is just because we take 'false belief' to be analogous with 'incorrect belief'.

moment, he wanted a cigarette, and if his belief were true he would have been able to get one. That much is true, but my aim is just to show here is just to point out that false beliefs in some contexts can have a higher utility than true belief all things considered, and my claim is that at least sometimes, such as in this case, the satisfying a higher-order belief at the expense of not satisfying a more immediate belief can be more beneficial to the individual. To help demonstrate this, we can easily imagine that the smoker is glad in retrospect that he was unable to find his cigarettes.

Finally, false beliefs can indirectly influence our behaviour such as to bring about positive consequences. Let's begin by considering a hypothetical of how this can come about. Imagine the case of a tribe that live close to a dangerous swamp. The swamp is dangerous because crocodiles live there that can attack and kill humans if they come close. The members of the tribe, however, don't fully understand the nature of the beasts that attack them, and put it down to the presence of evil spirits that live in the swamp. As a result, they avoid the swamp, and so avoid the danger. In this example, the false belief that *there are evil spirits in the swamp* is at least as beneficial as the true belief that *there are crocodiles in the swamp*; and it may even be more beneficial to the tribe if the false belief is more likely to keep members of the tribe away from the swamp than the true belief—because, say, evil spirits are more terrifying than crocodiles. So, false beliefs can be at least as beneficial as true beliefs by indirectly affecting action.

Moreover, we need not restrict ourselves to hypotheticals to make this point, as empirical evidence in fact supports that such cases do, in fact, occur. In particular, we can again turn to the literature on adaptive misbeliefs (introduced in Section 1.2). As I previously mentioned, one persuasive candidate for adaptive misbeliefs are positive illusions, in which case having overly positive beliefs about e.g. our own abilities can actually benefit our mental and physical health, even though they are false. However, there are also many

other potential candidates for adaptive misbeliefs, and Wilson and Lynn (2009) go so far as to say that they are ‘pervasive’. One example, defended at length by David Sloan Wilson (2010), is that religious beliefs are a form of adaptive misbelief. If true, this example sits nicely alongside the hypothetical swamp case above, even though that case is highly simplified compared to reality. The idea is that religious beliefs have adaptive benefits for individuals and societies despite being false,⁸³ such as promoting social cohesion through shared values etc. Adaptive misbeliefs such as these, therefore, are cases of false beliefs whose effects are beneficial to the individuals (and societies) that hold them.

Given that we are defining belief according to its effects, then, and false beliefs may have no effects (when they don’t perform their function) and can even have beneficial effects, I conclude that there is nothing essentially defective about false beliefs. For these reasons, my reading of doxastic correctness is thin, in that we can come to agree that true beliefs are often, or even mostly, the most useful beliefs to have; but there is no essential link between belief and truth such that false beliefs are substantively incorrect. This result is more palatable when we also reflect on the fact that the teleological and normative theories, which do give substantive readings of doxastic correctness, turn out to be problematic (as I argue in Part I and II respectively). Therefore, I hold, doxastic correctness is a pervasive, socially accepted construct. But beyond that it does not provide any information about an essential link between belief and truth.

14. Summary of ‘Functions’

In Part III, I turned my attention to answering the central questions of this thesis. I began by addressing Hume’s Problem. In answer to how beliefs can be distinguished from other forms of acceptance, I focused on the effects that beliefs have (or their outputs); this in

⁸³ Even if some religious beliefs are true, definitely most of them are false, given that different religious contradict each other.

in contrast to the teleological and normative theories of belief which focus on the causes of beliefs (of their inputs). By focusing on beliefs effects, I argued that we can distinguish beliefs from other acceptances according to the conditions contained in (what I call) the doxastic effects thesis.

Breaking this thesis down, it is constituted by two separate theses the conditions of which are all necessary for acceptances to count as beliefs, and are jointly sufficient for acceptances to count as beliefs. In particular, the two theses are a developed version of the traditional motivational role thesis, and the fundamentality thesis. More specifically, the traditional motivational role thesis dictates that beliefs must be able to motivate action, where the motivational properties of belief are to *cause* and *rationalise* action; and the fundamentality thesis dictates that beliefs must occupy a fundamental role in our mental framework, such that they *inform* and *regulate* our other acceptances. My claim, therefore, is that these conditions together, when they apply to an acceptance, make that acceptance a belief—thus they allow us to answer Hume’s Problem.

Following this, I went on to answer the second question: why we have beliefs (or, alternatively, why we have acceptances that satisfy these conditions). To answer this question, I endeavoured to interpret the components of the doxastic effects thesis as *functional statements*. To reach this end, I argued for a systemic theory of functions over the etiological theory, and then I applied the systemic theory to the motivational role and fundamentality theses. Once complete, the systemic analysis gives us an insight into why we have beliefs. In essence, the point is that beliefs, due to their motivational and fundamental *functions* cause higher-level effects up a systemic hierarchy. This in turn helps human populations to remain stable in an environment, such that the individuals in those populations who possess beliefs are able to reproduce their own kind. As such, the reason

we have beliefs is because beliefs, in virtue of performing their functions, are in part causally responsible for the reproduction of more individuals with beliefs.

Finally, continuing to focus on the effects of beliefs, I offered a thin reading of doxastic correctness, such that there is no essential link between belief and truth that makes true beliefs correct and false beliefs incorrect, but that instead holds that the principle has arisen as a social construct. This position becomes more acceptable once we realise that we are implicitly aware of motivational thesis; and in particular, the beneficial effects of having true beliefs in relation to satisfying our desires. This commitment gives rise to the idea that true beliefs are correct, in particular because usually we want to satisfy our desires, and we assume others are too when debating what to believe in the public domain. However, there are plenty of reasons for thinking that there is nothing essentially defective about false beliefs, especially when they either do not relate to action at all, or when they have beneficial effects of their own, such as when they help us to satisfy higher-order desires or when they are adaptive (as in, adaptive misbeliefs). For these reasons, and given the fact that I rejected the alternative (substantive) readings on doxastic correctness in Parts I and II, a thin reading of doxastic correctness should be preferred, and we should not think that false beliefs are essentially problematic (or in the substantive sense, essentially incorrect).

At this point, I have answered the three central questions of this thesis. But despite this, one important concern remains. Typically, theories of belief not only set out to provide conditions for *outright belief* and *outright disbelief*, as we have focused solely on so far, but also for a third doxastic state; namely, suspended belief (suspension). In the final part of this thesis, I ask one additional question: how do the conditions of the theories of belief discussed throughout, including my own, apply to suspension?

PART IV
SUSPENDED BELIEF

Part IV: Suspended Belief

15. Theories of Suspended Belief

Many theories of belief focus exclusively on outright belief and disbelief, however there is at least one other important doxastic attitude in our conceptual repertoire: suspension. Often, when consider whether to believe that p , we may not think we are in a position to form an outright belief, and so find ourselves in a state of suspension towards p . As such, we should expect theories of belief to extend to suspension as a doxastic attitude, in the sense that the conditions that account for beliefs also, in some sense, account for suspension. As we shall see, this is a point at least some theorists have recognised, and so they have attempted to account for suspension according to their theories of beliefs. But before moving on to assess the plausibility of these theories, let me begin by saying a few more words about *why* suspension is a doxastic attitude.

We might initially think, for instance, that suspension is not really an attitude at all, but the absence of an attitude. Here the term *withheld belief* comes to mind as a synonym for suspension. The idea of withholding from belief, more so than suspending belief, gives the impression that suspension is not an attitude, but is rather the lack of an attitude. By analogy, if we withhold an action, then we don't take any action; likewise, if we withhold a belief, then we don't form any belief. If this is true, then there is little sense in asking whether a theory of belief's conditions apply to suspension *as* a doxastic attitude. However, this doesn't seem quite right. Suspension is an attitude—which is a point Jane Friedman (2013) makes with a number of suggestive examples. For instance, are you suspending belief about whether I have blonde hair? Presumably, if you have never seen me before, you were not suspending belief about the colour of my hair—you'd never even thought

about.⁸⁴ And by the same token: ‘We don’t come into the world agnostic or suspending judgment about where bumblebees hibernate during winter’ (p. 168).⁸⁵ So, we don’t use suspension to refer to the mere absence of attitudes, it is an attitude that has something like the role of belief. Instead of believing or disbelieving a proposition, we can suspend belief about that proposition. For these reasons, I continue under the assumption that suspension is a doxastic attitude, and not the mere absence of an attitude. As such, conditions contained in theories of belief should not only account for outright belief and disbelief, but also suspension.

With this in mind, I now turn to each of the theories that have been discussed above, and ask how their conditions apply to suspension as a doxastic attitude. I argue that the ways of accounting for suspension that have been offered by the teleologists and the normativists are problematic, while the doxastic effects thesis proposes conditions that can account for suspension.⁸⁶

15.1. Against a Teleological Reading of Suspension

Recall that central to teleological theories of belief is the thesis that beliefs aim, in a descriptive sense, at truth. According to the truth-aim thesis, in conscious deliberation, we intend to believe that p only if p ; or when we don’t consciously deliberate, our subconscious mechanisms regulate for p only if p . Now, the important thing to note is that neither of these conditions apply directly to suspension, such that we can just straightforwardly count suspension as a doxastic attitude. We do not, that is, intend to suspend belief towards p only if p , and suspension is not regulated for p only if p . Rather,

⁸⁴ I’ve altered the content of one of Friedman’s examples, but the sentiment is the same.

⁸⁵ I use ‘suspension’ synonymously with Friedman’s ‘agnostic’ and ‘suspended judgment’.

⁸⁶ In this context, Tony Booth (2015) draws our attention to the fact that teleological and normative theories of belief have problems when it comes to account for suspension as an *involuntary attitude*. The arguments Booth proposes, if correct, are reason enough for denying that the teleologists and normativists can account for suspension. However, my aim is to block some further routes to suspension from these theorists.

suspension occurs precisely when we are *not* in a position to hold a doxastic attitude the content of which we take to be true. If we were in such a position, and we did take the content of our suspension to be true, then we would not be suspending about p , we would just have an outright belief that p . We don't suspend towards p , that is to say, only if p . So what can the teleologist make of suspension according to the truth-aim hypothesis? The solution I want to consider is suggested by Ernest Sosa (2008), who proposes that suspension is the result of a second-order aim that we have in relation to the first-order truth-aim.

To develop his position, Sosa makes use of many analogies between performing intentional actions and forming beliefs. He does so because he explicitly takes believing to be a kind of performance that is akin to intentional action. To get us thinking about suspension, one of the analogies he draws is to a hunter stalking his prey. The hunter, he suggests, has the primary aim to hit his target. This is analogous to the believer's aim to believe that p only if p (i.e. the believer's primary aim to hit the target involved in believing). However, the hunter also has a second-order aim, which is to take the shot at the appropriate time, such that when he does, he is confident that he can satisfy his primary aim to hit to the target. This second-order aim arises because the hunter doesn't just continuously keep shooting in the hope of hitting a target, instead he want to take the right shot at the right time. As such, the hunter has two aims:

(i) to hit the target,

and

(ii) to shoot only if he thinks he will hit the target.

As Sosa notes, the second-order aim can cause the hunter to 'intentionally... and deliberately forbear' from taking a shot, so as to 'avoid failure' concerning his primary aim (p. 11). In this way, the second-order aim of the hunter is derivative from the primary aim,

because he attempts to satisfy the second-order aim as a means of not violating the primary aim—which is where the truth-aim and suspension enter into the analogy.

In the process of deliberation, according to Sosa, we have the second-order aim to avoid failing to satisfy the (primary) truth-aim. But in this case, the second-order aim involves not going ahead and forming a belief that might be false. For this reason, when we cannot determine from the evidence whether to believe that p or to believe that not- p , we do not engage the primary aim of belief, to believe p only if p , and instead satisfy the second-order aim, to avoid failure, by deliberately forbearing from forming a belief. And, the point is, in the context of doxastic deliberation, this deliberate forbearance from belief, as a consequence of the second-order aim, is suspension.⁸⁷

The problem with this account of suspension, however, is that it fails to account for suspension *as an attitude* over and above the mere absence of an attitude. As discussed in the previous section, suspension is an attitude taken towards a mere proposition. It doesn't make sense to talk of individuals who have never had a proposition cross their mind as being in a state of suspension towards that proposition. Yet, Sosa's account doesn't make suspension anything more than the mere absence of belief. Suppose that we do, as Sosa claims, have the second-order aim to avoid failing to satisfy the truth-aim. In cases of doxastic deliberation, when we cannot determine the truth of either p or of not- p , we can satisfy the second-order aim by simply not forming any attitude at all towards p . By not forming any attitude, we succeed in not violating the truth-aim. So, the question is, where does suspension *really* enter into the analogy that Sosa gives? I don't see that it does. If our second-order aim is avoiding to have false beliefs, then we don't need suspension as a way of satisfying that aim. Having no belief, or no attitude at all, does the same work. As such, from Sosa's perspective, suspension has no role to play *as an attitude*.

⁸⁷ For his discussion and more of his examples see, Sosa (2008, 2010).

Hence, Sosa's development of the truth-aim to include a second-order aim does not account for suspension, so we are still left wondering what the teleologists can say about suspension as a doxastic attitude.

15.2. Against a Normative Reading of Suspension

As we saw earlier, normative theories of belief hold that beliefs are essentially subject to a truth-norm. This norm dictates that we ought to believe that p only if p .⁸⁸ Regarding suspension, the initial problem is that the truth-norm, so understood, never requires us to suspend belief. Indeed, it is never permissible to do so if we assume, as we are doing, that suspension is a doxastic attitude. This is a point Mayo (1963) notices against early versions of normativism:

[It follows from the truth-norm that] one ought never to suspend judgment, since there is nothing that it is right *not* to believe, except the false, the negation of which, being true, one ought to believe. (p. 144)

So, the normativists are struck by an immediate obstacle when it comes to accounting for suspension. Their primary thesis does not, on first sight, does not apply to suspension, and thus does not make suspension a doxastic attitude. As with the teleologists, then, we can ask what the normativists propose to do about this problem. One potential development, which I consider here, is offered by Engel (2013c).

To resolve the issue, Engel appeals to an evidential-norm, which is dependent on the truth-norm. He writes:

It is false that the norm of truth allows only two doxastic attitudes. If one considers whether p is true, and does not have enough evidence for either p or $\text{not-}p$, the norm does not prescribe believing p or believing $\text{not-}p$. It prescribes withholding belief. But isn't

⁸⁸ Another possibility is that the truth-norm states a permission: that we may believe that p only if p . For continuity, however, I continue to talk about the norm as a requirement. For discussion, see Section 4.1.

withholding belief then under the governance of the evidential norm...? It is, but it is also under the governance of the truth norm, for there is no possibility of being governed by the truth norm unless one follows the evidential norm. (p. 213)

In this passage, we see that Engel accepts there is some concern about how the truth-norm can account directly for suspension (or withheld belief, in his terminology). But he argues that we can reach suspension via an appeal to an evidential-norm for belief, which is 'under the governance' of the truth-norm. The idea is that like the truth-norm, the evidential-norm is an requirement, but this time to believe according to our (subjective) evidence, to the effect of: S ought to believe that p only if p is supported by evidence. This leads to suspension because, off the back of the evidential-norm, when we have insufficient evidence to hold an outright belief that p or that not- p , we are required to suspend belief about p . Nonetheless, it is difficult to see exactly how this is going to work as Engel intends, for a couple of reasons.

First, Engel runs into a problem similar to Sosa's—if there is an evidential-norm, it doesn't distinguish suspension from the mere absence of belief. To see why, we need to remember that the evidential-norm is in the service, so to speak, of the truth-norm. That is, we ought to believe according to our evidence ultimately because we ought to believe the truth, and evidence indicates truth. However, supposing this is true, the evidential-norm does not require us to suspend. For, if evidence is our indication of truth, and so we must believe the evidence to believe the truth, then when the evidence for a proposition is insufficient to determine the truth of that proposition, there is no reason for us, on the evidential-norm, to form any attitude at all. All that follows from the evidential-norm, as a norm derivative of the truth-norm, is that we ought to form outright beliefs in propositions for which we have sufficient evidence. By way of analogy: if you are subject to a normative requirement in a ball game, such that you ought to kick the ball only if you think you can score, and the only way to kick the ball is to swing your leg at it, then you are *not* required

to swing your leg at the ball if you don't think you can score, even if that is the only means of kicking the ball to score—at least, you are not required to do so according to the 'kick-the-ball-norm' alone. Similarly, even if believing the evidence is what we ought to do in order to satisfy the truth-norm, it doesn't then follow that we ought also to form intermediate attitudes, such as suspension, when we have insufficient evidence for an outright belief. So, even if we accept that the truth-norm entails the evidential-norm, and that we ought to believe the evidence in order to satisfy the truth-norm, we are still given no reason to hold an attitude—such as suspension—towards a proposition for which we have no evidence. In fact, all we need to do in the face of insufficient evidence for a belief is still hold no attitude at all. Thus, we see that Engel gives no role to suspension as distinct from the absence of having a doxastic attitude.

And second, we have no reason to agree with Engel in the first place that the truth-norm entails the evidential-norm. For we can satisfy the requirements of the truth-norm without satisfying those of the evidential-norm. Imagine for instance that your optimistic friend sincerely tells you that he is going to win the lottery this weekend. Given that the chances of him doing so are incredibly low, you ask him why he believes it. In response, he tells you that he knows the evidence suggest that he won't win, yet he believes it nonetheless. Then, as it happens, he goes and wins the lotter that weekend. In this case, your friend's belief was true, so he satisfied the truth-norm; but in holding his belief he completely contradicted what he knew was strong evidence against it. So, he did not satisfy the truth-norm via his commitment to the evidential-norm, and the evidential-norm is not entailed by the truth-norm.

For these reasons, the normativists, like the teleologists, still need to account for how their theory can account for suspension as a doxastic attitude. Next I consider how the essential properties of the doxastic effects theory apply to suspension.

15.3. A Doxastic Effects Theory of Suspension

As we saw in Section 7, the doxastic effects thesis is constituted by two separate theses: the motivational role thesis and the fundamentality thesis. These theses in turn require beliefs to possess four essential properties: that they can (i) cause action, (ii) rationalise action, (iii) inform other acceptances, and (iv) regulate other acceptances. The first two are the motivational properties, and the second two are the fundamental properties. I now argue that suspension, as is required of a doxastic attitude, possesses these properties.

15.3.1. The Motivational Properties of Suspension

According to the motivational role thesis, beliefs cause and rationalise actions. They *cause* actions alongside desires; and they *rationalise* actions (also alongside desires) in the sense that we explain our actions in terms of them. For instance, your belief that there is milk at the shop (in part) causes you to go to the shop, and if asked why you went to the shop, assuming you are being honest, you use the belief as part of your explanation of why you went to the shop—you went there because you thought they had milk (and you wanted milk). These properties are part of what makes an acceptance, then, a doxastic attitude.

The important thing here is that suspension also has these properties. And we can demonstrate this by use of an example. Imagine that it's now 7pm, and you want to get to the shop and back before 8pm. Usually, it takes you twenty minutes to get to the shop and back, and having a tendency to put things off until the last minute, you would typically set off at 7.40pm. However, you're not sure whether you will bump into a friend on the way, who will want to talk for a while; you don't believe it will happen, but you don't disbelieve it either—so, you suspend belief about whether you will bump into a friend. As such, given that your main priority is to get back before 8pm, you decide to set off

immediately, ensuring that if you do bump into your friend, you will be able to talk *and* get back in time.

In this case, your attitude of suspension about whether you will bump into a friend is responsible for you setting off earlier than you usually would. That is, it is one of the causes of your action. If you weren't suspending about whether you will bump into a friend—say you never thought about it—you would enact your normal procedure and set off later. Moreover, the suspension rationalises your action. Suppose you are asked why you set off early, by someone who know that you would usually go to the shop. One of the reasons you could give to explain your action is that you were suspending belief about whether you would bump into anyone, and so set off early just in case. Therefore, suspension can perform the motivational functions typical of outright beliefs; that is, suspension shares with belief the two essential motivational properties: to cause and rationalise actions.

15.3.2. The Fundamentality Properties of Suspension

The fundamentality thesis dictates that beliefs must be able to inform other acceptances. They do so by providing the information for the content of those acceptances. For example, when the child imagines that he is an elephant, his imagining that elephants have trunks is informed by his belief about elephants. In addition to the motivational properties, this fundamentality property is what makes an acceptance a doxastic attitude.

The important thing is that suspension also possesses this property—which we can see by means of an example, concerning how suspension informs other acceptances. Suppose that the child, again imagining he is an elephant, has an attitude of suspension about whether elephants can swim. As such, assuming also that he wants his imagining to be fairly realistic, he avoids bringing into his episode of imagining a situation where he, as an elephant, has to swim. In this way, suspension informs what the child imagines. In

contrast, things could be different if he had an outright belief about whether elephants can swim. If he believed that they can, he might be willing to bring into his imagining a situation where he pretends to be an elephant swimming; and if he believed they cannot, he might include a situation in which he is an elephant, say, drowning. But since he suspends about whether elephants swim, in order to keep his imaginings realistic, he avoids any scenario involving elephants swimming. So, suspended beliefs can occupy the role of outright beliefs insofar as they can inform other acceptances (such as imaginings).

Given the considerations in Section 15.3, then, we are able to say that suspension is a doxastic attitude on the doxastic effects theory. This is because suspension, like outright belief, is able to perform the motivational and fundamentality functions essential to being a doxastic attitude. Specifically, suspended beliefs can cause and rationalise actions, and they can inform other acceptances.

16. Summary of ‘Suspended Belief’

In this final part of the thesis, I extended the question about what beliefs are to a further doxastic attitude: suspension. Reflecting again on the theories of belief that we discussed throughout the thesis, I asked how those theories could apply to suspension to grant suspended beliefs doxastic status.

The motivating assumption, shared by many theorists, is that suspended belief is not the mere absence of an attitude, but is an actual doxastic attitude alongside outright belief and disbelief. Given this assumption, theories of belief should be able to account for suspension *as* a doxastic attitude. In this regard, I looked at some attempts by teleological and normative theorists to extend their theses to include suspension.

Despite Sosa’s appeal to second-order aims involved in the act of believing, I argued that the teleological theory fails to account for suspension. The main concern is that, even if

we have a second-order aim to avoid holding false beliefs, we have no reason to suspend (above and beyond just holding no doxastic attitude) when the evidence for an outright belief is insufficient. Simply forming no attitude at all satisfies our aim not to hold false beliefs. And concerning Engel's defence of a normative theory of suspension, we saw that the problem is twofold. First, even if the truth-norm does entail an evidential-norm, such that we ought to believe according to our evidence, this norm does not in fact require us ever to suspend, and in particular it does not require us to do so because it is in service of the truth-norm. As such, when we have insufficient evidence for the truth of a proposition it follows that we ought not to form any doxastic attitude in that proposition; and when we have sufficient evidence, then we ought to believe that proposition—there is no role for suspension distinct from the role of not forming any attitude at all. Thus, the normative account of suspension has an analogous problem to the teleological account. It doesn't make suspension anything more than the absence of belief, which conflicts with our assumption that suspension is an attitude. And second, it's not clear that the truth-norm entails an evidential-norm in the first place. If someone forms a truth belief, such as in the lottery case, even though they know it is not supported by the evidence, then they are abiding by the truth-norm yet rejecting the evidential-norm.

Finally, I argued that the properties appealed to in the doxastic effects theory do apply to suspension, and hence do account for suspension as a doxastic attitude. Suspension has the potential to be motivational, in the sense that it can cause and rationalise actions alongside beliefs (and desires); and it can occupy a fundamental role in our mental framework, by informing other forms of acceptances, again just like outright beliefs.

Given the points made in this discussion, I conclude that the doxastic effects theory succeeds where the teleological and normative theories fail: it accounts for suspension, as

well as outright belief and disbelief, as a genuine doxastic attitude. This is another reason for accepting, overall, my functional theory of belief over the alternatives.

Conclusion

Many philosophers have offered theories of belief. In this thesis my aim has been to explore these theories and provide a theory of my own, by focusing on three central questions: what beliefs are (Hume's Problem); why we have them; and what we should make of doxastic correctness. I interpreted the first question as a matter of how we can distinguish beliefs from other forms of acceptance, such as assuming, guessing, and (propositional imagining); the second question is about why we, as believers, have beliefs; and the third examines the relation between belief and truth, insofar as true beliefs are considered to be correct, and false beliefs incorrect.

In answer to these questions, I considered in Part I the plausibility of teleological theories of belief. Central to the teleological theories I discussed is the thesis that beliefs aim, in a descriptive sense, at the truth. As it turns out, this central thesis is problematic for various reasons. In particular, beliefs do not necessarily, and in the relevant sense, aim at the truth; the truth-aim thesis fails to distinguish beliefs from other acceptances; and interpreting beliefs as a teleological notion gives rise Owens' exclusivity objection—which, recall, is the objection that the aim of belief cannot be weighed against other aim-motivated behaviours. For these reasons, we cannot accept that teleological theories of belief provide the resources to either distinguish beliefs from other forms of acceptance, or to provide a plausible account of doxastic correctness. Furthermore, besides allusions to the fact that we have beliefs due to evolutionary processes, the teleologists do not go far enough in explaining why it is that we possess beliefs.

Following this, in Part II, I ask the same questions of normative theories of belief. The central aspect of normative theories is that beliefs are subject to a unique normative

requirement, such that the only appropriate thing to believe is the truth. Like the truth-aim thesis, the suggested truth-norm turns out to be problematic. First, it's not clear how the truth-norm is to be formulated, as every proposed formulation so far runs into problems. And second, because of our inability to believe, in conscious deliberation, anything other than (what we take) to be true, the truth-norm can't be genuinely normative. Unlike other norms, we cannot take any guidance from the truth-norm; and as we saw, this problem holds whether the norm is interpreted as a prescription, or, as Engel suggests, an idealisation. As such, the truth-norm thesis does not provide a satisfactory theory of belief. Given the problems it faces, we cannot use it to distinguish beliefs from other acceptances, or to account for doxastic correctness. In addition, the normativists make no clear efforts to explain why we have come to have acceptances that are subject to the truth-norm; that is, they fail to explain why we have beliefs.

In Part III, I then began trying to provide a theory of belief that does account for the three central questions. To do so, I focused on the effects (or outputs) of beliefs, as opposed to their causes (or inputs). This point represents a fundamental distinction between my theory of belief and the teleological and normative theories. In this way, I was able to begin providing answers to the central questions.

In response to Hume's Problem, I developed (what I called) the doxastic effects thesis. Constitutive of this thesis is a version of the traditional motivational role theory of belief and the fundamentality thesis. Together, these theses provide conditions for distinguishing beliefs from other forms of acceptance, by dictating the properties that beliefs must have in order to be beliefs. Specifically, beliefs must be able to cause and rationalise actions (the motivational role thesis) and inform and regulate other acceptances (the fundamentality thesis. If an acceptance possess each of these properties, then we can count it as a belief. Thus, the motivational role and fundamentality theses provide

necessary and jointly sufficient conditions for answering Hume's Problem. With a resolution to this problem at hand, I then turned my attention to explaining why we have beliefs. To do so, I turned to the literature in philosophy of biology on functions, and I defended a systemic theory of functions. This theory I then applied to the components of the doxastic effects thesis, thus interpreting the roles that beliefs have as functions. By doing this, I was able to provide an account that goes some way to explaining why we have beliefs: the functions of beliefs make a causal contribution, up a systemic hierarchy, to the stability of human populations, and this stability ensures that further individuals in that society, with the capacity for belief, reproduce. Finally, I considered what we could make of doxastic correctness from this proposed theory of belief. I suggested that doxastic correctness should be given a thin (as opposed to substantive) reading, such that we take beliefs to be correct as a social construct, but that there is nothing about beliefs that makes true beliefs essentially correct and false beliefs essentially defective. I then explore reasons for thinking that false beliefs can be just as useful to us as true beliefs, and in some cases even more so.

In Part IV I extended the question about what beliefs are to cover a third doxastic attitude; namely, suspension. I argued that various attempts to develop the teleological and normative theories of belief to account for suspension fail, but that my theory succeeds. In particular, the properties that the doxastic effects thesis dictates beliefs must possess, and that are sufficient for making an acceptance a doxastic attitude, also apply to suspension. That is to say, suspended beliefs share their ability to cause and rationalise actions, and to inform and regulate other acceptances, with outright beliefs.

For the many reasons discussed throughout, therefore, I conclude that a functional theory of belief is to be preferred to the teleological or normative alternatives.

Bibliography

- Alston, William P. (1988). The deontological conception of epistemic justification. *Philosophical Perspectives* 2:257-299.
- Amundson, Ron & Lauder, George V. (1994). Function without purpose. *Biology and Philosophy* 9 (4):443-469.
- Archer, Sophie. (2015). Defending Exclusivity. *Philosophy and Phenomenological Research* 93 (2).
- Armstrong, D. M. (1973). *Belief, Truth and Knowledge*. Cambridge University Press.
- Bennett, Jonathan (1990). Why Is Belief Involuntary? *Analysis* 50 (2):87 - 107.
- Bock, W. J., & Von Wahlert, G. (1965). Adaptation and the form-function complex. *Evolution*, 269-299.
- Boghossian, Paul A. (1989). The rule-following considerations. *Mind* 98 (392):507-49.
- (2003). The normativity of content. *Philosophical Issues* 13 (1):31-45.
- Boorse, Christopher (1976). Wright on functions. *Philosophical Review* 85 (1):70-86.
- Booth, Tony (2015). Belief is Contingently Involuntary. *Ratio* 29 (4).
- Buller, David J. (1998). Etiological theories of function: A geographical survey. *Biology and Philosophy* 13 (4):505-527.
- Bykvist, Krister & Hattiangadi, Anandi (2007). Does thought imply ought? *Analysis* 67 (296):277–285.
- (2013). Belief, Truth, and Blindspots In Timothy Chan (ed.), *The Aim of Belief*. Oxford University Press
- Chan, Timothy (ed.) (2013). *The Aim of Belief*. Oxford University Press.
- Côté-Bouchard, Charles (2016). Can the aim of belief ground epistemic normativity? *Philosophical Studies* 173 (12):3181-3198.
- Cummins, Robert C. (1975). Functional analysis. *Journal of Philosophy* 72 (November):741-64.
- (1983). *The Nature of Psychological Explanation*. MIT Press.

Davidson, Donald (1963). Actions, reasons, and causes. *Journal of Philosophy* 60 (23):685-700.

- (1987). Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association* 60 (3):441-458.

Davies, Paul Sheldon (2000). Malfunctions. *Biology and Philosophy* 15 (1):19-38.

- (2001). *Norms of nature: Naturalism and the nature of functions*. MIT Press.

Engel, Pascal (2005). Truth and the aim of Belief. In D. Gillies, (ed.), *Laws and Models in Science*. London: King's College Publications.

- (2007). Belief and normativity. *Disputatio* 2 (23):179-203.
- (2013a). In Defence of Normativism about the Aim of Belief. In Timothy Chan (ed.), *The Aim of Belief*. Oxford University Press
- (2013b). Sosa on the normativity of belief. *Philosophical Studies* 166 (3):617-624.
- (2013c). Doxastic Correctness. *Aristotelian Society Supplementary Volume* 87 (1):199-216.

Evans, Gareth (1982). *Varieties of Reference*. Oxford University Press.

Feldman, Richard (2001). Voluntary belief and epistemic evaluation. In Matthias Steup (ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Oxford University Press. pp. 77--92.

Frankish, Keith (2007). Deciding to Believe Again. *Mind* 116 (463):523 - 547.

Friedman, Jane (2013). Suspended judgment. *Philosophical Studies* 162 (2):165-181.

Gibbard, Allan F. (2003). Thoughts and norms. *Philosophical Issues* 13 (1):83-98.

- (2005). Truth and correct belief. *Philosophical Issues* 15 (1):338–350.

Glüer, Kathrin & Wikforss, Åsa (2009). Against Content Normativity. *Mind* 118 (469):31-70.

- (2010). The Truth Norm and Guidance: a Reply to Steglich-Petersen. *Mind* 119 (475):757-761.
- (2013). Against Belief Normativity. In Timothy Chan (ed.), *The Aim of Belief*. Oxford University Press.

- (2015). Still No Guidance: Reply to Steglich-Petersen. *Theoria* 81 (3):272-279.
- Griffiths, A. Phillips (1962). On Belief. *Proceedings of the Aristotelian Society* 63:167-186
- Hume, David (1739/2000). *A Treatise of Human Nature*. Oxford University Press.
- Kind, Amy (2001). Putting the image back in imagination. *Philosophy and Phenomenological Research* 62 (1):85-110.
- Kitcher, Philip (1993). Function and Design. *Midwest Studies in Philosophy* 18 (1):379-397.
- Kornblith, Hilary (2001). Epistemic obligation and the possibility of internalism. In Abrol Fairweather & Linda Trinkaus Zagzebski (eds.), *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*. Oxford University Press 231--248.
- Mayo, Bernard (1963). Belief and Constraint. *Proceedings of the Aristotelian Society* 64:139 - 156.
- McHugh, Conor (2011). What Do We Aim At When We Believe? *Dialectica* 65 (3):369-392.
- (2012). Belief and aims. *Philosophical Studies* 160 (3):425-439.
 - (2013). The Illusion of Exclusivity. *European Journal of Philosophy* 23 (4):1117-1136.
 - (2015). Engel on doxastic correctness. *Synthese*:1-12.
- McHugh, Conor & Whiting, Daniel (2014). The Normativity of Belief. *Analysis* 74 (4):698-713.
- McKay, Ryan T. & Dennett, Daniel C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences* 32 (6):493.
- Millikan, Ruth G. (1984). *Language, Thought and Other Biological Categories*. MIT Press.
- (1989). In defense of proper functions. *Philosophy of Science* 56 (June):288-302.
 - (1993). *White Queen Psychology and Other Essays for Alice*. MIT Press.
- Moran, Richard A. (1988). Making up your mind: Self-interpretation and self-constitution. *Ratio* 1 (2):135-51.

- (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press.

Myers (2002). *Social psychology*. 7th edition. McGraw-Hill.

Neander, Karen (1991a). Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science* 58 (2):168-184.

- (1991b). The teleological notion of 'function'. *Australasian Journal of Philosophy* 69 (4):454 – 468.
- (1995). Misrepresenting and malfunctioning. *Philosophical Studies* 79 (2):109-41.
- (2007). Biological Approaches to Mental Representation. In Mohan Matthen & Christopher Stephens (eds.), *Philosophy of Biology*. Elsevier.

Nichols, Shaun & Stich, Stephen P. (2003). *Mindreading. An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford University Press.

O'Brien, Lucy F. (2005). Imagination and the motivational role of belief. *Analysis* 65 (285):55-62.

Owens, David J. (2003). Does belief have an aim? *Philosophical Studies* 115 (3):283-305.

Papineau, David (1993). *Philosophical Naturalism*. Blackwell.

- (1999). Normativity and judgment. *Proceedings of the Aristotelian Society* 73 (73):16-43.
- (2013). There Are No Norms of Belief. In T. Chan (ed.), *The Aim of Belief*.

Plantinga, Alvin (1993). *Warrant: The Current Debate*. Oxford University Press.

Price, Carolyn (1995). Functional explanations and natural norms. *Ratio* 8 (2):143-160.

- (2001). *Functions in Mind: A Theory of Intentional Content*. Oxford University Press.

Railton, Peter (2003). Facts, Values, and Norms. *Philosophical Studies* 126 (3):449-462.

Ramsey, Frank Plumpton (1931). *The Foundations of Mathematics and Other Logical Essays*. Edited by R.B. Braithwaite, with a Pref. By G.E. Moore. --. Routledge and Kegan Paul.

Shah, Nishi (2002). Clearing Space For Doxastic Voluntarism. *The Monist* 85 (3):436-445.

- (2003). How truth governs belief. *Philosophical Review* 112 (4):447-482.
- Shah, Nishi & David Velleman, J. (2005). Doxastic deliberation. *Philosophical Review* 114 (4):497-534.
- Sinhababu, Neil (2013). Distinguishing Belief and Imagination. *Pacific Philosophical Quarterly* 94 (2):152-165.
- Sober, Elliott (1984). *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. University of Chicago Press.
- Sorensen, Roy A. (1988). *Blindspots*. Oxford University Press.
- Sosa, Ernest (2003). The Place of Truth in Epistemology. In Linda Zagzebski & Michael DePaul (eds.), *Intellectual Virtue: Perspectives From Ethics and Epistemology*. New York: Oxford University Press. pp. 155-180.
- (2008). Knowing full well: the normativity of beliefs as performances. *Philosophical Studies* 142 (1):5-15.
 - (2010). *Knowing Full Well*. Princeton University Press.
- Steglich-Petersen, Asbjørn (2006). No Norm needed: On the aim of belief. *Philosophical Quarterly* 56 (225):499–516.
- (2009). Weighing the aim of belief. *Philosophical Studies* 145 (3):395-405.
 - (2017). Weighing the Aim of Belief Again. *Logos and Episteme* 8 (1):141-145.
- Stich, Stephen P. & Nichols, Shaun (2000). A cognitive theory of pretense. *Cognition* 74 (2):115-147.
- Sullivan-Bissett, Ema (2014). *Belief, Truth, and Biological Function* (Ph.D. Dissertation). University of York.
- (2016). Malfunction Defended. *Synthese*:1-22.
- Sullivan-Bissett, Ema & Noordhof, Paul (2013). A defence of Owens' exclusivity objection to beliefs having aims. *Philosophical Studies* 163 (2):453-457.
- Sylvan, Kurt 2016. The illusion of discretion. *Synthese* 193 (6):1635-1665.
- Taylor & Brown 1988. Illusion and Well-Being: A Social Psychological Perspective on Mental Health. *Psychological Bulletin*. 103(2): 193-210.

- Vahid, Hamid (2006). Aiming at Truth: Doxastic vs. Epistemic Goals. *Philosophical Studies* 131 (2):303-335.
- Van Leeuwen, Neil (2009). The Motivational Role of Belief. *Philosophical Papers* 38 (2):219 - 246.
- Velleman, David (2000). On the aim of belief. In *The Possibility of Practical Reason*. Oxford University Press. pp. 244--81.
- Walton, Kendall L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press.
- Wedgwood, Ralph (2002). The aim of belief. *Philosophical Perspectives* 36 (s16):267-97.
- (2007). *The Nature of Normativity*. Oxford University Press.
 - (2013a). The Right Thing to Believe. In Timothy Chan (ed.), *The Aim of Belief*. Oxford University Press. pp. 123-139.
 - (2013b). Doxastic Correctness. *Aristotelian Society Supplementary Volume* 87 (1):199-216.
 - (2017). Replies: on norms of belief and knowledge. *Synthese* 194 (5):1555-1564.
- Whiting, Daniel (2010). Should I Believe the Truth? *Dialectica* 64 (2):213-224.
- (2013a). Nothing but the Truth: On the Norms and Aims of Belief. In Timothy Chan (ed.), *The Aim of Belief*. Oxford University Press.
 - (2013b). Truth: the Aim and Norm of Belief. *Teorema: International Journal of Philosophy* 32 (3):121-136.
- Whyte, J. T. (1990). Success Semantics. *Analysis* 50 (3):149 - 157.
- (1997). Success again: Replies to Brandom and Godfrey-Smith. *Analysis* 57 (1):84--88.
- Williams, B. (1970). Deciding to believe. In Bernard Williams (ed.), *Problems of the Self*. Cambridge University Press:136-51.
- Wilson, D.S., 2010. *Darwin's cathedral: Evolution, religion, and the nature of society*. University of Chicago Press.

Wilson, David Sloan and Lynn, Steven Jay (2009). Adaptive misbeliefs are pervasive, but the case for positive illusions is weak. In Ryan T. McKay & Daniel C. Dennett. The evolution of misbelief. *Behavioral and Brain Sciences* 32 (6):493.

Wright, Larry (1973). Functions. *Philosophical Review* 82 (2):139-168.

- (1976). *Teleological Explanations: An Etiological Analysis of Goals and Functions*. University of California Press.