

# The Unruly Mind

## Against Doxastic Normativism

Thesis for the degree of Doctor of Philosophy  
Word count 74,815

Johanna Magdalena Schnurr  
St Hilda's College, University of Oxford  
Trinity Term 2017





## Abstract

This thesis is an evaluation of doxastic normativism. Doxastic normativism is the theory that belief is essentially normative, and that the norm of belief is a constitutive norm. After introducing the taxonomy of normativist views I use throughout the thesis, I discuss four versions of doxastic normativism: state normativism, conceptual normativism, intentional teleologism, and teleo-functionalism. The first two are those traditionally referred to as “doxastic normativism”, and they are the main targets of the thesis. Discussion of the second two is less extensive, and serves to show that the central problem for normativism lies not in the kind of normativity that it claims is essential to belief, but in the attempt to make any normative property whatsoever a necessary condition for belief. I develop undercutting objections to two arguments for state normativism and to one argument for conceptual normativism. My objections draw on i.a. the semantics of correctness ascriptions, the norms of other intentional states, the nature of rationality, and the evidential sensitivity of belief. Then there are several rebutting objections, drawing on non-agential mental states, the normative classification of epistemically deviant mental states, and difficulties for state-differentiation through norms. The discussions of teleologism and teleo-functionalism demonstrate that these views, although making much weaker claims about belief, fall prey as well to the main problems identified for the core normativist views. The thesis closes on a chapter offering reflections about the reasons behind the failures of normativism, as well as a descriptive theory of belief that I argue can fulfil the explanatory desiderata of doxastic normativism.



for Mona



## Acknowledgements

I am indebted beyond anyone else to my supervisor, Timothy Williamson. My work has benefited from his formidable philosophical mind, but just as importantly I have benefited from his guidance and advice during various challenges. I was equally lucky in my secondary supervisor, Bill Child, who helped finetune the project in its final stages and thanks to whose scrutiny many of the ideas in it improved substantially. I received helpful comments at various stages of this research from Beau M. Mount, Mona Simion, Lidewij Arthers, Rachel Fraser, Daniel Waxman, Stephen Wright, and conference audiences in Leuven and Cambridge. For editing help, I am grateful to Tom Partridge, Emily Mayne, and Robert Hortle.

I owe particular thanks to my dear friend Beau M. Mount, for his brilliance and great company; to Mona Simion, for serving as a continuous inspiration and for her faith in me; to my brother Florian, for instrumental conversations about biology and for his rigorous writing advice; and to my parents, Annette and Otto, for their unconditional support.

## Table of Contents

|  |    |
|--|----|
| Introduction.....  | 11 |
| Chapter 1: Groundwork and Definitions .....                      | 15 |
| §1.1 Belief .....  | 16 |
| §1.1.1 Belief and other epistemically relevant states.....       | 16 |
| §1.1.2 The semantics of affirmative attitudes .....              | 20 |
| §1.1.3 Confidence, commitment, candidate-beliefs.....            | 21 |
| §1.2 Norms.....  | 23 |
| §1.2.1 Norms and their origin.....                               | 24 |
| §1.2.2 Normative scepticism .....                                | 25 |
| §1.3 Doxastic normativism .....                                  | 26 |
| §1.3.1 Normativism narrowly and widely construed .....           | 28 |
| §1.3.2 Essences and definitions.....                             | 29 |
| §1.3.3 Kinds of doxastic normativism .....                       | 32 |
| §1.3.4 Intrinsic norms .....                                     | 34 |
| §1.3.5 Normative thickness.....                                  | 37 |
| §1.4 Other stipulations.....                                     | 39 |
| Chapter 2: Arguments for State Normativism .....                 | 41 |
| §2.1 The norm of narrow normativism.....                         | 41 |
| §2.1.1 Instrumental, categorical, and axiological norms .....    | 41 |
| §2.1.2 State and conceptual normativism.....                     | 43 |
| §2.2 The argument from correctness conditions.....               | 44 |
| §2.2.1 The extended argument.....                                | 45 |
| §2.2.2 Belief's correctness condition .....                      | 49 |
| §2.2.3 Correctness and normativity .....                         | 51 |
| §2.2.4 The semantics of correctness ascription .....             | 52 |
| §2.2.5 More ambiguity.....                                       | 58 |
| §2.3 The dilemma of correctness conditions.....                  | 60 |
| § 2.4 The argument from dispositions.....                        | 64 |
| §2.4.1 Essential dispositions .....                              | 65 |
| §2.4.2 Resisting the normativity of rational dispositions .....  | 66 |
| §2.4.3 Resisting the rationality of essential dispositions ..... | 69 |
| §2.4.4 Irrationality of defeater-insensitivity .....             | 71 |
| §2.5 Resisting state normativism.....                            | 72 |
| Chapter 3: Against State Normativism .....                       | 75 |
| §3.1 Non-agential mental states .....                            | 75 |
| §3.1.1 Problems for non-agential attitudes .....                 | 76 |
| §3.1.2 The argument from vestigiality.....                       | 81 |
| §3.1.3 The norms of proto-beliefs.....                           | 84 |
| §3.2 Dispositions and overdiscriminate definitions.....          | 87 |

|  |     |
|--|-----|
| §3.3 Norm guidance.....  | 89  |
| §3.3.1 Content normativism and doxastic normativism .....          | 89  |
| §3.3.2 Intrinsic norms and guidance.....                           | 91  |
| §3.4 The lawless deviance objection.....                           | 95  |
| §3.4.1 Believers and beliefs .....                                 | 96  |
| §3.4.2 Degrees of guidedness .....                                 | 97  |
| §3.4.4 A minimal threshold?.....                                   | 101 |
| §3.4.5 Lawless deviance.....                                       | 102 |
| §3.4.7 State-differentiation .....                                 | 105 |
| §3.5 Against state normativism .....                               | 108 |
| Chapter 4: Against Conceptual Normativism .....                    | 111 |
| §4.1 The argument from transparency.....                           | 112 |
| §4.1.1 Teleologism and transparency.....                           | 113 |
| §4.1.2 Normativism and transparency .....                          | 116 |
| §4.2 Against the argument from transparency .....                  | 118 |
| §4.2.1 Is deliberation about whether to believe transparent? ..... | 118 |
| §4.2.2 Is deliberative belief uniquely transparent?.....           | 123 |
| §4.2.3 Does the normativist account work?.....                     | 124 |
| §4.3 Against conceptual normativism.....                           | 128 |
| §4.3.1 Conceptual disagreement.....                                | 128 |
| §4.3.2 Lawless deviance.....                                       | 130 |
| §4.3.3 State-differentiation .....                                 | 133 |
| §4.3.4 Psycho-ontological indistinguishability.....                | 134 |
| §4.3.5 Normativism without correctness .....                       | 137 |
| §4.4 Non-normative state-differentiation.....                      | 143 |
| §4.4.1 Belief and action .....                                     | 145 |
| §4.4.2 Exceptions to strong Belief-Action.....                     | 147 |
| §4.4.3 Action-inspired normativism .....                           | 152 |
| §4.4.4 The function of weak affirmative attitudes .....            | 153 |
| §4.6 Against conceptual normativism.....                           | 155 |
| Chapter 5: Teleological definitions of belief.....                 | 159 |
| §5.1 What kind of aim? .....                                       | 160 |
| §5.1.1 Aim effectiveness.....                                      | 162 |
| §5.1.2 Norm tempering.....   | 164 |
| §5.1.3 Telos by extension.....                                     | 166 |
| §5.2 Instrumental belief norms .....                               | 170 |
| §5.2.1 Optional normativity.....                                   | 172 |
| §5.2.2 Derived instrumental norms against lawless deviance.....    | 173 |
| §5.3 Against intentional teleologism.....                          | 175 |
| Chapter 6: Teleo-functional definitions of belief.....             | 177 |
| §6.1 Functions and norms.....                                      | 177 |

|  |     |
|--|-----|
| §6.2 Preliminaries of etiological teleo-functionalism .....                  | 180 |
| §6.2.1 Functions and functioning.....  | 181 |
| §6.2.2 Biological items .....  | 183 |
| §6.3 Belief as function-constituted .....                                    | 187 |
| §6.3.1 Functional monism.....  | 188 |
| §6.3.2 Conflicting functions.....  | 190 |
| §6.3.3 Conjunctive functions? .....  | 192 |
| §6.3.4 Shifting functions .....  | 195 |
| §6.3.5 The functions of belief.....  | 198 |
| §6.4 Against teleo-functionalism.....  | 201 |
| Chapter 7: Towards a Knowledge-First Theory of Belief .....                  | 203 |
| §7.1 Differentiating Belief .....  | 203 |
| §7.1.1 Belief and Knowledge.....   | 204 |
| §7.1.2 Knowledge-compatibility.....  | 205 |
| §7.1.3 Knowledge-compatibility and conceptual analysis .....                 | 210 |
| §7.2 Knowledge-compatibility and the psychology of belief .....              | 211 |
| §7.2.1 Doxastic involuntarism .....  | 212 |
| §7.2.2 Kinds of freedom .....  | 214 |
| §7.2.3 Ignorance as failure.....   | 217 |
| §7.2.4 Belief and dispositional ignorance-ascription.....                    | 220 |
| §7.2.5 Freedom from affirmative attitudes.....                               | 223 |
| §7.3 Normative differences between belief and other affirmative states ..... | 226 |
| §7.4 The failures of state-centrism.....                                     | 229 |
| §7.4.1 Over- and underprescription .....                                     | 229 |
| §7.4.2 State-overarching epistemic normativity .....                         | 233 |
| §7.4.3 Overdiscriminate state-differentiation .....                          | 234 |
| §7.4.4 Normativity and psychology.....                                       | 237 |
| Conclusion.....  | 241 |
| Bibliography .....   | 251 |
| Index.....   | 259 |

## Introduction

It matters what we believe. Beliefs determine how we behave, feel, think; they make us who we are. But while almost everybody agrees on the importance of getting our beliefs right, it is less obvious what it is that makes our beliefs right: what should they have in common in order to be the kinds of beliefs we ought to have? Philosophers treat this question, and related ones, under the topic of the *norm of belief*. Doxastic normativists, or normativists about belief, have an answer to the question how our beliefs ought to be, and they think there is a good reason for that answer. Because what makes our beliefs good beliefs, according to the normativists, is tied up in the very nature of what belief is. Indeed, they think there is no way to really understand what belief is without understanding what its norm is. Just like what makes something a (tennis) serve has to do with what makes for a good tennis serve, and what makes something a promise is determined by what makes for a good promise: their normative aspects are bound up in their nature. Normativism about tennis serves seems like a good approach: a serve would not be what it is if the rules that determine what makes for a good serve were different, hence incorporating those rules in our theory of serves makes sense. And indeed, when we look up "serve" we tend to find the rules governing them explained prominently, usually before discussion of technique.

But promises and tennis serves have in common, along with games and other speech acts the study of which has inspired doxastic normativism, that they are practices, socially constructed, cultural kinds. Is belief like them? When we call belief a mental state, is it a state for the mind to be in in virtue of entirely descriptive facts, the way a physical object is in a given temperature state entirely in virtue of descriptive facts? Or is it a state to be in in virtue of normative facts as

well, the way a football player is in off-side position not just because of his physical location but because of the rules governing his physical location as well? If you are like me, you find the former picture more appealing: of belief, qua state for the mind to be in, as a natural kind, like boiling is a natural kind qua state for a liquid to be in. But this thesis won't argue that belief is a natural kind, although that may be supported by its conclusions. It will argue that the opposing view, of belief as normatively constituted, cannot be right. It cannot be right for two broad categories of reasons. Firstly, because it entails problematic things about the nature of belief. Specifically, the way it distinguishes beliefs from things that are not beliefs is marred by serious difficulties. Secondly, because it entails problematic things about the norm of belief. The kind of belief norm that we can justify through normativism is not the kind of norm we want for belief. It condemns things it shouldn't and it fails to condemn things it should.

This is a thesis against normativism about belief but it is not a thesis against belief norms. Indeed, much of what I will argue takes as a premiss that the norm of belief should be a strong norm: a norm that issues genuine obligation and genuine censure and that cannot be opted out of. A norm that is rooted in the nature of the mind cannot deliver on those demands. The mind is too unruly to be its own lawbook.

There are four versions of doxastic normativism I will engage with in detail. The thesis is largely structured around those discussions. The first chapter explains what I understand by doxastic normativism, and what its four instantiations have in common. It also features other definitions of concepts important to the thesis, specifications of terminology, background assumptions, and clarifications of theoretical commitments. The following five chapters all engage with specific versions of normativism, with the discussion of state normativism cut into two chapters,

the first presenting undercutting objections to arguments that have been made for the view, the second presenting rebutting objections. Then there is another very substantial discussion of conceptual normativism, and two shorter chapters that engage with teleologism and teleo-functionalism respectively. After these critical discussions, I finish on a chapter presenting a positive account of belief and epistemic normativity.

Despite the classification of the central five chapters by sub-view, they at times feature lengthy discussions of problems for normativism generally. Most of the main objections I will present hold against multiple or all formulations of normativism, but the scope of this thesis will not allow me to spell them out multiple times. Instead I limit myself to remarks about how a certain problem could be brought to bear against other sub-views, and revisit these ideas in the conclusion which attempts to explain how the many arguments against versions of normativism add up to a rejection of doxastic normativism *tout court*.

The seven chapters are divided into titled sections numbered with two-place labels (§X.X), many of which are further divided into titled subsections with three-place labels (§X.X.X). I will cross-reference to sections and subsections in the thesis by § numbers. At the end of the thesis there is an index with definitions, thought experiments, and important terms for easy reference.



## Chapter 1: Groundwork and Definitions

### Belief as an affirmative attitude, norms, species of normativism

Doxastic normativism is a theory of belief that defines it through appeal to normative criteria. As such, it aims to provide answers to two kinds of questions: what should we believe, and what are our beliefs like? The first of these questions has its home predominantly in the literature on epistemic norms. What normative constraints does epistemology put on us? What are our obligations as rational agents? Applied to our mental life specifically, this yields questions like: what ought we to believe? How does rationality dictate we should think? Other areas of investigation of epistemic norms are norms of assertion and norms of practical reasoning, results from which have inspired work on the norm of belief in places.

Despite being sometimes referred to as the “ethics of belief”, the question what we should believe is usually seen as one relevant to and answerable by epistemology. Morality, of course, has bearing on what we should believe. If an evil mind reader tells me they will murder five innocent people for every type of fruit that I believe raccoons can eat, that datum will certainly influence what I should believe about raccoon diet. However, moral considerations about what will kill the least innocent people are not naturally read as telling us what is rational to believe, or which beliefs will get us knowledge, or maximise epistemic utility, or are faultless, or justified. It is epistemically normative evaluations like these we usually have in mind when we ask what we should believe, and it is these that are up for debate in the literature on the norm of belief.

The idea that there is a specifically epistemic brand of normativity<sup>1</sup> is shared by most contributors to the literature, with David Papineau (2013) providing a well-known exception. In this first chapter, I will focus on the basic notions we need to evaluate doxastic normativism, with belief and norms being the foremost ones. I will give definitions of the specific views we will encounter throughout the thesis and explain what they have in common.

## §1.1 Belief

Belief is a propositional attitude. It is a mental state directed at a propositional content, and it is *about* that proposition. This aboutness makes it a member of the large class of intentional mental states. Other examples of intentional attitudes are judging that *p*, assuming that *p*, fearing that *p*.

### §1.1.1 Belief and other epistemically relevant states

Belief is a member of a specific subset of intentional states that is especially relevant to epistemology. These are attitudes like guessing, judging, assuming, being certain, suspecting, being inclined to believe, and conjecturing. They can be contrasted with intentional attitudes like desiring, hoping, fearing, imagining, or entertaining, which seem far less epistemically relevant. Many attempts have been made to capture the difference between these two groups. Firstly, guessing, judging, assuming, etc. are sometimes called “doxastic states” (Engel, 2013). The states in question certainly are belief-like in a sense but for our purposes the label “doxastic” is unfortunate, since much of the dissertation is devoted to finding out what exactly makes belief

---

<sup>1</sup> Which is irreducible to moral norms and whose demands can indeed be in conflict with them.

the kind of state it is, *within* the set of propositional attitudes to which it is most similar. Another label that comes up in the literature is “cognitive attitudes” (Velleman, 1992; Shah & Velleman, 2005), as contrasted with conative attitudes like desires<sup>2</sup>. But what makes cognitive and conative attitudes into the types of mental states they are respectively?

One early and still frequently appealed to notion to distinguish belief-like states is direction of fit. Originally found in the philosophy of language (Austin, 1953), the distinction between different directions of fit was introduced to the philosophy of mind with the terminology of “mind-to-world” and “world-to-mind” direction of fit by Searle (1983), who credits Anscombe (1957) for the idea. Mind-to-world direction of fit is supposed to characterise beliefs: their content is meant to fit the way the world is. World-to-mind direction of fit on the other hand applies to desires: they mean to make the world fit their content.

There is continued disagreement between descriptive and normative interpretations of direction of fit. Smith (1987), exemplary of the former along with Humberstone (1992), for instance, defines belief descriptively in terms of direction of fit: “[A] belief that p tends to go out of existence in the presence of a perception with the content not p, whereas a desire that p tends to endure, disposing the subject in that state to bring it about that p.” (Smith, 1987, p. 115). Against this notion of belief as characterised by how it *tends* to behave, authors such as Zangwill (1998) and Gregory (2012) argue that direction of fit captures a difference in how beliefs *ought* to behave with respect to the world, compared to how desires ought to behave

---

<sup>2</sup> The distinction between the cognitive and the conative has existed in empirical psychology for a long time (Ostrom, 1969).

with respect to the world. Others argue against the use of any formulation of direction of fit for the characterisation of these mental states (Frost, 2014).

On either the descriptive or the normative interpretation, direction of fit is a first step towards seeing what belief has in common with states like guessing and assuming. Both are distinct from conative attitudes, in that they are expected to fit the world and not to make the world fit them. Yet I will not rely on the direction of fit (DOF) framework to class belief with its intentional kin. There are several reasons for this. The first is that authors who have used DOF in their theories of belief, most notably Velleman, class imagination as having the same DOF as belief, and hence as one of the cognitive attitudes. But imagination does not share the element of epistemic relevance that unites belief, guessing, estimating etc. and is therefore crucially different from these states for reasons explored below. I argue that belief belongs to a class of attitudes that is subtly different to the set of attitudes that have been called cognitive or mind-to-world fitting. The second reason not to rely on the DOF framework stems from cases of belief with an ambiguous fit. There are some scenarios where wishful thinking seems designed to *make* its content true. Confidence-buffering beliefs along the lines of “I can do this” can, when they work well, make it more likely that the subject can indeed do it. If DOF is descriptive, it seems clear that these beliefs do not have the DOF we expect for beliefs.. Since we want to count these self-fulfilling states as beliefs, using descriptive DOF to characterise belief-related states is out of the question. But since my thesis is an evaluation of normative definitions of belief, starting out with

a normative definition of the class of attitudes that it belongs to is problematic. Besides, self-fulfilling beliefs can also make trouble for normative definitions of direction of fit.<sup>3</sup>

I gave some examples of attitudes that I consider related to belief in a way relevant to this research (guessing, assuming, estimating). What is it that unites these attitudes if not their direction of fit? The subset of intentional states I am interested in is characterised by representing a proposition and having an attitude towards it, but also by that attitude being one of taking a stance on the proposition's truth. When one believes, or judges, or guesses etc. that *p*, one is invested in *p*'s truth. For that reason, these states all share the property that one can be mistaken in them. I call these mental states affirmative attitudes.

Def. 1.1 affirmative attitudes: intentional mental states that involve an endorsing or affirmation of the truth of their propositional content, where said affirmation can come in various degrees of strength

The following is a (very plausibly non-exhaustive) list of affirmative attitudes:

|             |               |            |
|-------------|---------------|------------|
| Believing   | Guessing      | Judging    |
| Assuming    | Hypothesising | Supposing  |
| Speculating | Estimating    | Suspecting |

---

<sup>3</sup> It's not always clear that beliefs with a descriptively speaking world-to-mind DOF are normatively speaking mind-to-world (and thus in violation of the norms governing them). In the case below, it is at least somewhat plausible that the subject does the right thing with her belief *epistemically speaking*, in which case we might have to say that the belief in question has a world-to-mind DOF normatively speaking as well (on which direction it counts as fitting): Ex. 1.1: Christina is a biologist studying a virus. Faced with a big deadline, she finds herself needing to work long hours through yet another weekend. Feeling her motivation leave her, she convinces herself that the work she will do this weekend has a high probability of yielding results that will be instrumental in the development of a lifesaving vaccine. This isn't true and is also not born out by the evidence of the immediate impact Christina's weekend work is likely to have. But Christina doesn't look at the evidence, forms the relevant belief about how crucially important this work is, and thus inspired, she works hard through the weekend and meets her deadline for a paper that provides one more small step towards understanding the biology of the virus. After her successful collection of the data, she revises her belief about the probability of aiding in the discovery of the vaccine to whatever her evidence supports.

Being inclined to believe

Being tempted to believe

Surmising

Being certain

I call these states “affirmative” for several reasons. Firstly, they all involve a stance that is stronger than neutrality. In forming one of these attitudes, the subject in a sense says internally “yes” to their content, although with various levels of force and with differing phenomenology. Secondly, affirmative attitudes share the element of stance-taking with the speech act of affirmation. The speech act involves a public stance-taking, which is a form of commitment. Affirmative mental states are not public but they too involve the undertaking of a commitment: affirming a proposition gives one a stake in its truth. Of course, we mentally and publicly affirm propositions that we would much prefer to be false. But there is a clear sense in which, if we end up being lucky and the proposition with the dreaded content turns out false, it would have been better of us if we had not affirmed it.

### §1.1.2 The semantics of affirmative attitudes

It is crucial to point out that many of the words denoting affirmative attitudes are ambiguous and also denote things that are not affirmative attitudes. The word “assume” is a prime example. We can assume things for the sake of argument. Very often, especially amongst philosophers, this involves no endorsing of the content of the assumption at all. Indeed, we can assume things for the sake of argument that we are inclined to believe are false, or in a *reductio ad absurdum* even things we know to be false. This sense of assuming is a speech act as well as a mental operation. It is not an affirmative attitude. The following is an example of assuming qua affirmative attitude. My friend Alice is very often late. I’m about to meet Alice for coffee at 16:30, and I say to myself “There is a good chance Alice won’t get there until 16:45”. I arrive at 16:25

myself and bring a book. When Alice hasn't arrived yet at 16:35, I am not surprised. I had assumed she would be late.

The words "hypothesise" and "conjecture" are similarly ambiguous. They can firstly refer to a propositional attitude that involves affirmation, and secondly to a mental operation in which one takes no stance on the truth of the proposition one uses as a hypothesis or conjecture. In that second sense, the hypothesis is merely something one postulates to fulfil a certain functional role within an investigative context (like allowing one to design experiments, or allowing one to think about what would follow from the proposition if it were true). The word "guessing", too, displays some ambiguity: there is a notion of guessing as a move in game-like contexts or bets. Say I'm in the final round of a quiz and I have to answer a question with two possible answers between which my evidence is completely neutral. By picking the right one I can double my current winnings whereas by picking the wrong one I will merely lose ten percent off my current winnings. I will probably go ahead and "make a guess". Performing the action of guessing in that context does not require an affirmative stance on the answer one picks: it is compatible with an agnostic mental state.

Unless explicitly stipulated otherwise, when I speak of these mental states in this thesis, I have in mind only assumptions, hypotheses, conjectures and guesses *qua affirmative attitudes*. This semantic dimension is essential to bear in mind.

### §1.1.3 Confidence, commitment, candidate-beliefs

Then there is the question where credences fit among the affirmative attitudes. The answer is, they don't. Credences are *confidence* levels, but belief and the other affirmative attitudes as I conceive of them involve more than confidence: they involve *commitment* as well. To see the

difference between confidence and commitment, consider a lottery. Having bought a ticket to a lottery of which I know the number of tickets (1000), I adopt credence 0.999 towards the proposition that my ticket is a loser. My ticket turns out to win. Now qua confidence level, I clearly didn't make a mistake when I had credence 0.999: that correctly captured the epistemic probability. A belief that my ticket is a loser however would have been a mistake, no matter how favourable the probabilities of truth were. This difference in whether falsehood of the proposition marks a mistake is due to the affirmative element of belief: belief represents its content as true, credence represents its content as having a certain probability. One can also be mistaken in one's credences: had I adopted credence 0.7 or credence 1 in the lottery scenario above, that would have been a mistake *even if* my ticket had in fact been a loser. Belief consists of both confidence and commitment (and my hypothesis is that the same is true for the other affirmative attitudes). On that view, there may well be an answer to the old Lockean threshold question, as long as that question is understood as asking for a necessary condition on belief<sup>4</sup>, not a sufficient one: there may be a lower threshold of confidence below which an attitude cannot count as belief, but meeting that confidence threshold is not enough. The agent needs to endorse the content of her attitude in an affirmative way as well, thus undertaking a commitment.

Including the commitment element in my definition of belief (and the other affirmative attitudes) is crucial for the work I undertake in this thesis, because it underpins much of the kind of normativity I will discuss. For instance, the truth norm of belief that I engage with in many places would not fit well with a credence-reducing view of belief on which the appropriateness of the

---

<sup>4</sup> In fact, some of the ideas I will present in the final chapter have bearing on the confidence threshold of belief, although that will not be the reason I present them.

attitude is entirely a matter of adequate representation of epistemic probability. Although I will present many qualms with the truth norm, I share and endorse the central intuition that false beliefs are mistaken. Therefore, neither belief nor the other affirmative attitudes reduce to confidence levels on my view.<sup>5</sup>

Further relevant vocabulary concerns the classification of given token mental states into types. Since much of this thesis is concerned with the definition of belief, we will look at many counterexamples to proposed definitions from the literature. Often, these counterexamples will involve a case in which some subject's described token mental state is excluded from counting as a belief on the definition under evaluation, even though the case strikes us intuitively as involving a belief. In setting up these cases I will speak of *candidate-beliefs* as those states the doxastic status of which we are investigating.

## §1.2 Norms

At the heart of the debate about the norm of belief is the question what criterion or criteria our beliefs ought to meet in order to be epistemically how they ought to be. Usually, accounts of "the norm of belief" will formulate some answer or other to that question. I will refer to this as the question about the *content* of the norm of belief. I divide the content question up further into the criterion, the species and the form question. The first of these asks what *criterion* it is that gives beliefs their normative status. Candidate answers include truth, evidence, knowledge, wellbeing, and expected utility. The *species* question asks what kind of norm the belief norm is.

---

<sup>5</sup> In my definition of affirmative attitude, I mentioned only affirmation/commitment and not confidence. The reason for that is the assumption that one cannot affirm a content one does not have a certain minimal confidence about. If that assumption is mistaken, the definition should be updated to include this minimal confidence level as a necessary condition (where for some attitudes like guessing and assuming the only confidence condition may well be that one needs at least credence 0.5).

Examples of kinds of norms are deontic norms, axiological norms, functional norms, and instrumental norms. Deontic norms are norms that prescribe, forbid, and allow specific things. Axiological or evaluative norms assign things a normative property, like good, bad, justified, or correct. Functional norms are an important sub-species of axiological norms that evaluate when things function correctly. Instrumental norms are norms designed to help subjects fulfil their goals, and hence their force is conditional on what the subject aims to achieve. The *logical form* question asks how we should spell out the norm of belief: is it a conditional, a biconditional, does it have a wide or narrow scope, does it take beliefs or believers as its object?

### §1.2.1 Norms and their origin

Norms can be investigated along more dimensions than the content of what they tell us to do. Another such dimension is the *source* of the norm: where do its demands originate from? It seems quite straightforward, for instance, to say that norms of etiquette originate from convention and that traffic norms originate from encoded traffic law. Applied to the norm of belief, some candidate answers to the source question are the believer's aims and goals (the norm of belief stems from the aim of believing), their self-interest (the norm of belief stems from practical rationality), biology (the norm of belief stems from the biological function of our minds), morality (the norm of belief reduces to moral demands) and belief itself (the norm of belief is a part of the concept or state of belief).

Different combinations are possible between candidate answers to the content and source questions, but we tend to find some correlations in the literature. Firstly, if one favours a strong deontic belief norm, this will be more difficult to argue for with an aim or function as the source of the norm. An evaluative norm on the other hand, especially one with properties like success

as the criterion, goes particularly well with a teleological sourcing. Norms that evaluate cognitive mechanisms instead of the mental states they produce can be founded well in a functional framework, and the normative significance of evaluative criteria like reliability can also be accommodated well by an appeal to selectional histories. We can thus expect answers to the species question and answers to the source question to work together.

### §1.2.2 Normative scepticism

A related question to the source question is the *authority* question: by virtue of what does the norm of belief have an actual say over believers? Some norms, like the rules of a game, have to be opted into in some sense to have a say over us, others, like the norms of morality, are commonly understood as being impossible to opt out of. Can epistemic normativity be opted out of? Is the importance of living up to the norm of belief a function of how much being rational or knowledgeable matters to one? Negative answers to these questions are instrumental to the field of normative epistemology, since epistemic norms are meant to censure irrational, unfounded, false attitudes regardless of what the bad believer in question happens to find important or which norms they would like to apply to themselves. For if we allow for the norm of belief to be strongly goal-dependent, strictly optional, or sourced in arbitrary conventions or misapprehensions, normative scepticism looms. The sceptical answer to the authority question is simply that epistemic norms have no real authority, that they might be useful tools to help us form beliefs in accord with the facts or with our evidence if that's what we feel like doing, but that no one can censure us for "failing" to do so, or for "failing" to want to do so.

Inquiring into the content of the norm of belief might be an interesting activity among philosophers who all share the conviction that this norm is real and important, but without an account of its source that can provide an explanation of its authority, the resulting investigation will be of limited use. It should be clear that if we want to avoid scepticism about the authority of the norm of belief, it will be useful to have an answer to the source question: anyone disinclined to accept the authority of epistemic norms as brute fact can be pointed to their source to explain why they should accept said authority.<sup>6</sup> This is where doxastic normativism comes in.

### §1.3 Doxastic normativism

Doxastic normativism is an answer to the source question. The norm of belief, according to doxastic normativism, stems from the nature of belief itself. Here is a first approximation of a definition of doxastic normativism.

Def. 1.2 doxastic normativism: it is part of the essence of belief that it is subject to the norm of belief

This means that belief by definition has normative constituents: we cannot define belief without mentioning normative properties. It also means that the norm of belief applies to belief by definition and simply in virtue of what belief is. Therefore, anyone who has beliefs automatically falls under the norm. The authority question is thus answered in the desired way. This section

---

<sup>6</sup> The teleological answer to the source question, which as explained below I will count as an instance of doxastic normativism, might strike us as an instance of the kind of scepticism about norm authority that normativism was meant to avoid. But as will become clear in the discussion of teleologism in section 3.3 below and then at length in Chapter 5, despite the fact that teleologism answers the species question instrumentally and the source question with a theory about believers' aims, the possibility of opting out of the belief norm is excluded.

presents a brief overview of the normativist views that the thesis will address. The four views will be explained in much greater detail in the chapters devoted to them respectively.

Normativism's proposed connection between norm and nature can take on a number of forms. On the most straightforward conception, the norm of belief is a part of the nature of the mental state of belief. I call this view *state doxastic normativism*. Proponents include Robert Brandom, Kate Nolfi, and Ralph Wedgwood (Brandom, 2001; Nolfi, 2015; Wedgwood, 2002). They observe that it seems to be a part of the nature of beliefs, for instance, that they can be true or false. If this property of truth-aptness is interpreted as a normative one, a norm of belief like "believe only true things" can be derived from that.

Def. 1.3 state doxastic normativism: it is essential to the mental state belief that it is subject to the norm of belief

One consequence of state normativism is that epistemic accounts of the norm of belief will have bearing on the philosophy of mind. A highly relevant literature from that field is that on intentionality. What makes propositional attitudes the kind of mental states they are? How does their aboutness relate to the nature of the mind? Proponents of normative theories of intentionality take the norms that govern the content of propositional attitudes to be a marker of the intentional: the norms that govern them set intentional states apart from the rest of the world. Depending on whether that normativity is seen as something that can be explained in terms of, or brought back to, non-normative properties (the teleosemantic project is an example of an attempt to do that), the normativity of the intentional can even be construed as indicating the irreducibility of mind to nature.

Doxastic normativists can and do make use of this idea that the intentional is normative. As we will see, several of the authors who contribute to the epistemological investigation of what we ought to believe have commitments about the nature of intentionality as well (Glüer & Wikforss, 2009, 2013b; Papineau, 1999; Wedgwood, 2009). The relevance of candidate answers to the source question about the norm of belief to the philosophy of mind will be important to keep in mind, for as we will see some of the proposals of doxastic normativism carry consequences for what views of the mind are available to us.

On a proposal methodologically closely related to state doxastic normativism, belief's normative properties are not situated on the level of the mental state but on the level of our concept of belief. I call this view *conceptual doxastic normativism*. Proponents include Paul Boghossian, Pascal Engel, and Nishi Shah (Boghossian, 2003; Engel, 2004; Shah, 2003; Shah & Velleman, 2005). When we think about beliefs, it seems clear that it is part of our concept of them that they can be true or false. The conceptual normativists argue that we can interpret that element of the concept normatively, and hence a norm of belief follows.

Def. 1.4 conceptual doxastic normativism: it is essential to the concept belief that belief is subject to the norm of belief

### §1.3.1 Normativism narrowly and widely construed

At this stage, I need to introduce some terminological distinctions. In the contemporary literature on belief norms, when we see the word "normativism", it usually applies to state and/or conceptual doxastic normativism. These two are often contrasted with teleologism (discussed below) which is presented as a rival view. Yet I will treat teleologism as a kind of normativism. The same goes for teleo-functionalism about belief. The next two subsections justify that

approach. To avoid confusion, I will speak of *doxastic normativism narrowly construed* as the set of views that includes state and conceptual normativism, and that coincides with the view that is usually called “doxastic normativism” in the literature. I will speak of *doxastic normativism widely construed* as the set of views that includes not only state and conceptual normativism but also teleologism and teleo-functionalism. It is doxastic normativism widely construed that is the topic of this dissertation, and that is referred to in the title. I will make the distinction explicit repeatedly in this introduction, but for the rest of the thesis, where I speak of “doxastic normativism” without further qualification, it is its widely construed formulation that is at issue.

Def. 1.5 doxastic normativism narrowly construed: it is essential to belief that it is subject to the norm of belief

Def. 1.6 doxastic normativism widely construed: the essence of belief entails that it is subject to the norm of belief

These definitions draw on the notion of essence. I will not have the opportunity to defend an account of essence here, but will rely on Kit Fine’s work to explain what I have in mind.

### §1.3.2 Essences and definitions

In a highly influential paper (1994), Fine starts out with the observation that our intuitive understanding of essence does not reduce to modal properties. The essence of  $\Phi$  does not coincide with everything that is necessarily the case as a consequence of  $\Phi$ ’s existence. This is one of his examples to that effect:

Consider two objects whose natures are unconnected, say Socrates and the Eiffel Tower. Then it is necessary that Socrates and the Tower be distinct. But it is not essential to Socrates that he be distinct from the Tower; for there is nothing in his nature which connects him in any special way to it. (p. 5)

For that reason, essence is not thought to be closed under entailment: there are things that logically follow from something's essence, that are not part of that essence. This is what marks the difference between narrow and wide normativism. On both accounts, the norm of belief follows from the essence of belief. Yet on narrow normativism, it follows from the essence because it is part of that essence, whereas wide normativism is compatible with the norm *merely* following from the essence. For the views that I will call merely widely construed normativist (teleologism and teleo-functionalism), this might be conservatively construed. Take the example of intentional teleologism about belief, according to which it is essential to belief that it aims at the truth. From that essence, it follows that belief is subject to an instrumental belief norm. But perhaps we want to go further and say its being subject to instrumental normativity is itself essential on teleologism.<sup>7</sup> In that case, the view would count as not merely widely normativist. But one upshot of Fine's insight about the closure failure of essences is that demarcating essences is not straightforward. I do not want my classification of the targeted views as species of normativism to be jeopardised, so I make use of the distinction between being part of belief's essence and following from belief's essence, even though that may ultimately not be necessary.

In follow-up work to "Essence and Modality", Fine introduced the distinction between *consequential* and *constitutive* essence (Fine, 1995). Constitutive essence is that which an object has essentially but not because it is a consequence of some more basic essential property. Consequential essence, which is the kind usually referred to by the word "essence" is not closed under entailment. Consequential essence however, is: everything that follows from an object's essential properties is its consequential essence. To take Fine's example, it is part of the

---

<sup>7</sup> Being able to account for belief norms is certainly a major explanatory desideratum for teleologism.

consequential essence of Socrates that he is a human or an amphibian, but only his being a man (and not that property's logical consequences such as disjunctions) is part of his constitutive essence. Thus, there is another way to make the distinction between wide and narrow normativism. According to narrow normativism, normative elements are part of the constitutive essence of belief, whereas according to wide normativism they are part of the consequential essence of belief. But again, it is difficult to delineate where constitutive essence ends and merely consequential essence begins.

Constitutive or consequential, we will see that the connection between belief and its norm is construed as a much more intimate one than the connection between Socrates and his being either human or amphibian, on all versions of normativism we will study. Before we turn to them, a word on definitions. Gideon Rosen (2015) explains well why, even on a definition of definition that appeals to essences, definitions are not the same as essences:

[I]t's worth stressing that these notions are not equivalent. On Fine's account, absolutely everything has an essence. Given any item  $x$ , definable or not, there will be truths of the form  $\Box x p$ , and we may identify  $x$ 's essence with the class of such truths. Even if the Gettier examples show that knowledge is indefinable (Williamson 2002), it still lies in the nature of knowledge that if  $S$  knows that  $p$  then  $p$  is true. Similarly, while negation is almost certainly indefinable, it presumably lies in the nature of negation that  $p \rightarrow \sim\sim p$ . Such examples show that Fine's notion of essence is different from, and more general than, the concept of real definition we've been discussing. (p. 194)

The literature I am engaging with is heavily invested in definitions of belief. It is not incoherent though to stipulate that belief has a normative essence (even in the constitutive sense) whilst refusing to offer a definition of belief that includes that normative element, for instance because one holds that belief cannot be defined. I will speak throughout the thesis of faults with certain definitions of belief. But at the root of these problems is not the attempt to completely define

belief. One major group of problems has to do with the attempt to use normative conditions as a necessary condition on belief. Another has to do with attempts to use normative conditions as a sufficient condition to differentiate belief from other affirmative attitudes<sup>8</sup>. When I argue against certain definitions of belief, my objections are not (primarily) due to the potential undefinability of belief. Even qua principles about belief that are not offered as part of a complete definition, the normativist conditions fail.

### §1.3.3 Kinds of doxastic normativism

State and conceptual doxastic normativism as defined above are species of doxastic normativism narrowly construed: they (partially) define belief through the norm of belief, which is said to be essential to it. Qua narrowly construed normativist views, they are of course also widely construed normativist views. The other two normativist (widely construed) views we will evaluate are teleologism and teleo-functionalism. These theories explain the origin of the norm of belief in a more roundabout way than normativist proposals, involving an extra inferential step. However, they ultimately too entail that belief has normative properties. Looking at the actual accounts will make all of this clearer. *Teleologism* defines belief as a state that by definition involves certain aims of the subject, with the aim at truth being the most prominent proposal. This teleological property makes it so that anyone who believes has a reason to want to believe truly, since this is an aim of theirs simply by virtue of believing.

Def. 1.8 teleologism: belief is the cognitive attitude a subject has if and only if it is their aim to hold that attitude only if its content is true

---

<sup>8</sup> Just because some property of belief suffices to distinguish it from guesses and assumptions doesn't mean that property is a sufficient condition for belief. Flight distinguishes bats from other mammals, but it is not a sufficient condition for something being a bat.

Aims necessarily generate instrumental norms. If I aim to steal your Picasso, then there is a sense in which I ought to steal your Picasso, in light of which if I do not steal your Picasso I will have *failed*. If a subject aims to mentally represent p only if p is true, then there is an instrumental norm according to which she ought to represent p only if p is true. And since that aim is part of what it is to be a believer, the norm of belief applies to all beliefs.

Def. 1.9 teleological norm of belief: the norm of belief is instrumental to the aims any believer has in virtue of believing

Teleologism makes the notion that belief has an aim concrete by linking it to the believer's intentions. Another way of filling in the metaphor of the aim of belief is by appealing to the concept of functions. One could say, for example, that the function of the mental state belief is to represent things as they are. One well-established tradition links functions to evolutionary histories that explain why the biological entity in question exists. This is a version of *teleo-functionalism* about belief, on which it is its selectional history that makes a biological entity the kind of thing it is.

Def. 1.10 teleo-functionalism: belief is the attitude that has representing true content as its proper function

A thing's proper function provides the conditions for its correct functioning. Something having a proper function therefore generates evaluative norms to the effect of whether the thing is functioning correctly or not.

Def. 1.11 teleo-functional norm of belief: the norm of belief is an evaluation of whether the belief-producing processes function in accordance with the task they were selected for producing

#### §1.3.4 Intrinsic norms

Doxastic normativism widely construed involves an explanation of the source of the norm of belief: it is entailed by its essence. This also involves a specific answer to the species question though: if the norm of belief is sourced in belief, that means that it is an *intrinsic* norm. Intrinsic norms apply to tokens of a type they range over by virtue of what it is to be a token of that type.

Take the following, uncontroversial, norm for (non-purely-ornamental) vases: vases ought to be water retentive. That a vase should be able to retain water (not be porous, not have holes, not be soluble) seems intrinsic to what a vase is. Now compare this with the following hypothetical norm for vases: vases ought to be purple. Say we are in a country with a dictator who has a strong hatred for vases that are any other colour than purple. The purple-vase norm is therefore decreed by law. Say that this dictator will also kill five people for every non-purple vase that is made. The purple-vase norm is therefore a moral norm. If the five people in question include the maker of the offensive vase, we could say that it is a prudential norm as well. Under none of these guises (legal, moral, prudential), is the purple-vase norm intrinsic to what a vase is or to what vase-making is. If the dictator is usurped tomorrow by his sister who has a similarly intense preference for yellow vases, the purple-vase norm will be replaced by a yellow-vase norm, but the nature of vases will not change.

The intrinsic/extrinsic distinction applies to the candidates for the source of the norm of belief we looked at above. On a proposal where the norm of belief reduces, for instance, to one's (non-belief-specific) moral obligations (e.g. allow no harm), the norm of belief will be extrinsic

to belief and to believing. The same is the case for proposals on which the norm of belief reduces to self-interest.

Proposals about the norm of belief could be construed either intrinsically or extrinsically. The teleologist norm of belief, for example, is an instrumental norm conditional on the subject's aims. If I aim at the truth in believing, it follows that my beliefs ought to be true in a sense. Now if aiming at the truth were something I do contingently, the norm of belief would not be intrinsic to belief: it would be possible for there to be times where my beliefs are not accompanied by the goal to believe the truth, and hence the goal-derived norm of belief would not apply to those beliefs. Extant defences of a teleological norm of belief however are made in the context of a teleological account of what belief is, and hence go further than the claim that the norm of belief is conditional on the subject's aims: said aims are *necessarily connected* to believing. The situation where one believes something in the absence of the aim to believe the truth is hence excluded. On this view, the norm of belief is in a sense intrinsic to belief, since aiming at the truth is intrinsic to the activity of believing and the belief norm follows from the aim at the truth.

On teleo-functionalism, the belief norm is an axiological norm that ascribes a normative property like success to beliefs that are produced by normally functioning belief-producing processes. Functions are intrinsic to the things they are the function of. It is part of the nature of vases that they are supposed to hold water. It is part of the nature of the thyroid gland that it is supposed to produce thyroid hormones. Teleo-functionalism entails that it is a part of the nature of whichever unit or process produces beliefs, that it has producing true beliefs as its

proper function. The evaluative norm of fulfilling that function or failing to fulfil it is then again tied to belief intrinsically.

Teleologism and functionalism make the belief norm intrinsic to belief via certain detours. For teleologism, it is a detour via the aims of the believer which the teleologist links to belief as a matter of constitutive essence. For functionalism, it is a detour via the function of the belief-producing processes. Belief is stipulated to necessarily stem from certain processes, and those processes are stipulated to necessarily have a certain function. Doxastic normativism narrowly construed makes the belief norm intrinsic to belief in a more direct way. State normativism makes the norm of belief a part of the nature of belief: beliefs just are those things that are governed by the norm of belief. Conceptual normativism makes the norm of belief a part of our concept of belief: when we think about belief, we think about it by definition as governed by the norm of belief.

For my purposes, teleologism, functionalism, state normativism and conceptual normativism are all examples of views that propose a norm of belief intrinsic to belief. This is another way to characterise them as species of doxastic normativism widely construed. We can see the intrinsicness these candidate belief-norms have in common if we compare them to proposals on which the belief norm is extrinsic. Since normativist norms are intrinsic norms, without them, belief would not be what it is. In contrast, if the belief truth norm "believe only what is true" were derived from the moral norm "do no harm" in conjunction with certain facts about the world that make it so that all false beliefs do harm, that norm would be extrinsic to belief. For if the world changes radically tomorrow, so that all true beliefs do harm instead, and the truth

norm therefore ceases<sup>9</sup> to be, it does not follow that beliefs will be different kinds of mental states in this new world. On the other hand, if it were no longer true that one furthers one's belief-aims by believing truly, or that believing truly fulfils a certain function of one's cognitive apparatus, or that believing truly is in accord with the norm of belief that is a part of the nature of belief, this would have radical consequences for what beliefs are (indeed, whether they are) according to teleologists, functionalists, and narrow normativists.

### §1.3.5 Normative thickness

Intrinsicness marks a certain subset of norm species. Another distinction in norm species is that between deontic and axiological norms. One dimension of difference between the candidate belief norms of these different species is something I will call normative thickness. One way to understand the metaphor of norm thickness is through the amount of censure an agent will be subject to as a consequence of a norm applying to them. Prescriptive norms range over agents and tell them to do or not do certain things. Infringing on a prescriptive norm is a failure on behalf of the agent. Such a failure therefore makes the agent liable to be censured.

The notion of *liability to censure* I have in mind has little to do with the actual probability of an agent being criticised by another agent. There might be some beliefs one would never in fact receive criticism for, perhaps because it is impossible for others to find out one holds them, perhaps because, although wrong, everybody else holds the same belief, perhaps because no one cares about epistemic norms. In these circumstances, one can still be liable to censure if it would be plausible to have one's doxastic situation described in the third person and for

---

<sup>9</sup> It is one limitation of this thesis that it does not address what it means for norms to exist or for them to be true. I will use these interchangeably: if there exists a norm for belief that says that it ought to be true, then it is true that beliefs ought to be true.

someone hearing that description to assert that one should not have believed as one did in those circumstances. If I falsely and on very poor justification believe that there were more than 3000 books in my grandmother's study on the 18<sup>th</sup> of October 1952, I am liable to censure for that belief according to certain deontic belief norms, even if this belief in fact never ends up getting me into trouble. If my cat however fails to believe that the Pythagorean theorem is true, my cat is intuitively not liable to censure for this.

Evaluative norms can carry censure as well. "You did a bad thing" is not much less forceful qua criticism than "you did something you ought not to have done". However, some evaluations are more censuring than others. "Bad" censures quite heavily for beliefs, as does "irrational". A range of other evaluative predicates carry censure as well, but usually in a less damning way: "unwarranted" and "unreliable" are examples, as are evaluations from the functional realm, like "unsuccessful" or "functioning poorly".

Normativity thus comes in degrees in a sense. At the same time, the distinction between normative and non-normative judgments is a categorical one. Sorting matter into organic versus inorganic, evaluating numbers as even versus uneven, evaluating books as hardcover or paperback etc. are all completely non-normative judgments. Normative evaluations can follow from such categorisations. If I am only supposed to sell hardcover books, whether some book in my store is hardcover or paperback will be normatively significant. But even if the normative property "permissible to sell" neatly supervenes on the non-normative property of being hardcover, it is clear that being hardcover is still an entirely non-normative property. This is important to bear in mind for beliefs: it's clear that from "this belief is true", normative things follow in conjunction with other facts. If all judgments of the form "this belief is true" happen in

the same circumstances as judgments of the form “this belief is good” or “this is the belief the subject ought to have”, that still doesn’t make “true” a normative property.

#### §1.4 Other stipulations

Some further clarifications need to be made before we launch our investigation of doxastic normativism. The first concerns the concept *epistemic norm* itself. Current epistemological literature has not yielded a common definition of the epistemic. Saying that the epistemic consists of the set of those objects and properties that are relevant to knowledge, we have a good place to start. Belief, truth, evidence, specific kinds of evidence like testimony, memory, or perception, reliability, defeaters, rationality: they all fall within the set of things that are relevant to knowledge. Desire, beauty, unity, or happiness are more rarely relevant to what we know and how we know it, so the investigation of those objects is much less relevant to the study of knowledge. When we talk about *epistemic norms* however, matters are less straightforward. For one, there is an ongoing and intense disagreement between truth norm and knowledge norm defenders, for both belief as well as assertion. Calling the norm of belief an epistemic norm from the outset might thus be misleading if what is at issue is whether knowledge is the criterion for permissible belief. Throughout this thesis, when I speak of epistemic norms, truth norms are included, for “epistemic” is used not as synonymous with knowledge, but as the denominator for all that is relevant to the study thereof, which obviously includes truth. I will not develop or defend a system for typing norms into epistemic and otherwise. The norms we will encounter in this thesis are all, unless marked otherwise, epistemic. They are norms that regulate behaviour relevant to knowledge, such as belief, by appeal to permissibility or evaluation criteria relevant to knowledge, such as truth or evidence, and they

are designed to bring about states of affairs that are favourable to one's standing with regard to knowledge, like true belief or knowledge itself.

There are some theoretical commitments that are popular throughout the epistemic norms literature I am engaging with. One of them is *doxastic involuntarism*. Ever since Bernard Williams' (1970) work on the aim of belief, the presumed impossibility of believing at will has played in the background of the debate. Doxastic involuntarism is the thesis that doxastic voluntarism is false: we cannot directly influence our beliefs. This is often attributed to the idea that belief formation can take only perceived evidence as its input: we cannot form a given belief simply because we want that belief, we need to perceive some evidence for its truth. I will engage with different formulations of doxastic involuntarism at length in Chapter 7. My own view is that doxastic voluntarism may well be true, and that much of what would speak in favour of its truth or falsity is outstanding empirical evidence. However, I will not assume the truth of doxastic voluntarism in any of my arguments, and when investigating normativist arguments I will sometimes assume the truth of formulations of involuntarism that are much stronger than my own views so as to evaluate the arguments on their own terms.

A principle that is even more widely accepted in the literature than doxastic involuntarism is that *ought implies can*. Again, I will take this on board in places where it plays a role in someone's argument, but that does not necessarily imply an endorsement of the principle on my part, and none of my own arguments should depend on it.

## Chapter 2: Arguments for State Normativism

### Correctness conditions, dispositions to believe

State normativism is a species of doxastic normativism narrowly understood, which I defined as the thesis that belief has normative properties essentially, and thus that in order to define what belief is, we need to mention normative properties. This distinguishes it from the widely construed version of normativism, on which the norm of belief follows from belief's definition through things like the intentions of the believer or her evolutionary history, which are not normative in themselves. Yet instrumental norms follow from intentions and axiological norms follow from functions. The status of the belief norm is intrinsic on both kinds of normativism: it follows necessarily from the nature of belief, although with a different number of inferential steps.

#### §2.1 The norm of narrow normativism

On this schema, normativism narrowly understood could include the same species of norm as the wide varieties. Therefore, a view on which belief has instrumental or functional norms in its definition would count as normativist narrowly understood. In practice however, we see the defenders of normativism narrowly understood argue for species of belief norms with a much thicker normativity: their favoured norm is categorical instead of instrumental and usually deontic instead of axiological/functional.

##### §2.1.1 Instrumental, categorical, and axiological norms

One difference between categorical and instrumental norms is the kind of censure they open one up to. Criticising someone for violating an instrumental norm is always indexed to their

goal. Say I have the goal to buy duck eggs but I don't succeed and come home with chicken eggs instead. You say to me "You shouldn't have gone to Aldi if you wanted duck eggs". Without reference to my goal, there is no sense in which I shouldn't have gone to Aldi, or in which there is something wrong with showing up with chicken eggs. Categorical norms are different. A norm that says "S ought to  $\Phi$ " unconditional on any of S's interests, means that when S doesn't  $\Phi$ , she did something wrong, and we can tell her "You should not have  $\Psi$ d" or "You should have  $\Phi$ d" without reference to her interests. We will return in more detail to the nature of instrumental normativity and its relevance for belief in §5.2.

In the extant accounts of normativism narrowly understood that we will study, the norms are categorical instead of instrumental. They are also usually prescriptive<sup>10</sup>, which sets them apart from the evaluative norms generated, for instance, by teleo-functionalism (see Chapter 6). Categorical axiological norms do entail liability to public censure, although as I argued in §1.3.4, they come in different normative thicknesses (calling a belief bad seems more damning than calling it unsuccessful). The evaluative norms of teleo-functionalism however don't lead to censure by themselves, because they are norms for biological entities (belief-producing processes), not norms for agents. Agents can be censured for their malfunctioning cognitive apparatuses by the lights of functional norms only when we combine those functional norms with a norm for agents that says: "S ought to be such that the biological entities she is made up of function properly".

---

<sup>10</sup> With some exceptions (Fassio, 2011, 2016; McHugh, 2012b).

## §2.1.2 State and conceptual normativism

The extant versions of doxastic normativism narrowly construed promise us a norm that generates epistemic obligations<sup>11</sup> without the need for bridge premises or reference to the aims of the believers. This chapter is dedicated to the arguments that are meant to link such a strong belief norm constitutively to the nature of belief, and hence warrant a definition of belief states in normative terms. Some of the defenders of state normativism are Brandom (2001), Wedgwood (2002, 2007, 2009) and Nolfi (2015). The latter writes:

Normativism enjoys a great deal of popularity in epistemology because it promises to explain, in a relatively straightforward way, the authority of epistemic norms: on any version of normativism, beliefs will be appropriately evaluated with respect to epistemic norms simply by virtue of being the sorts of mental attitudes that they are. There are, however, surprisingly few<sup>12</sup> sustained defences of normativism about the nature of belief- only a handful of philosophers have tried to provide independent motivation for the view. (p. 181)

In this chapter, we look at two such existing arguments: the argument from correctness and the argument from dispositions for state doxastic normativism.

---

<sup>11</sup> Or, in the exceptions mentioned above, evaluations.

<sup>12</sup> Nolfi mentions Wedgwood (2007), Zangwill (1998, 2005, 2009), Shah (2003) and Shah and Velleman (2005) as authors who have offered positive arguments for normativism. Wedgwood, Zangwill and Nolfi herself all use versions of an argument from dispositions to believe, which is the type of argument I am attacking in §2.4. Shah and Shah and Velleman offer the argument from transparency. However, Shah and Shah and Velleman explicitly argue for normativism about the concept of belief. On their account, it is only those beliefs the subject consciously adopts after deliberation whether to believe *p* that are normatively constituted. This is because deliberating about whether to believe by definition invokes the concept of belief, which they argue is normative. Shah's and Shah and Velleman's account is compatible with the possibility that the nature of belief as a mental state is not normative, and I therefore reject Nolfi's classification of their argument from transparency as a "defence of normativism about the nature of belief".

## §2.2 The argument from correctness conditions

The argument from correctness is an extension of a line of reasoning we find throughout the literature on belief norms<sup>13</sup>: that we can derive the norm of belief, specifically its criterion, from belief's correctness condition. That the correctness condition is truth is near-universally (see Smithies (2012b) for a rare exception) accepted: "The following is, surely, the merest platitude: For any p, believing that p is correct if and only if it is true that p." (Whiting 2012, p. 286)

Correctness is normative according to the defenders of normativism, which is seen as similarly obvious as the first premiss. So we get:

- P1) A belief is correct iff it is true (CC)
- P2) Correctness is normative for belief
- C) It is a norm for belief that it is correct iff it is true

Much discussion then goes into the question what the normativity of correctness consists in: is it obliging, permissive, evaluative, or something else? Then we get:

- P3) The normative dimension of correctness is obligation/permission/positive evaluation
- C2): There is a norm of belief that you are obliged/permitted to believe p / that your belief that p gets positive evaluation if and only if it is true that p

Call this *the content argument from correctness*: from belief's correctness condition we get the criterion of the belief norm (truth), from an analysis of the concept correctness we get the species of the belief norm (deontic, axiological, etc.). Note that the content argument is not an argument for doxastic normativism, but merely an argument for a norm of belief with a certain content. In this sense it differs from patterns of arguments from correctness such as that which

---

<sup>13</sup> Including in arguments for conceptual normativism.

Glüer and Wikforss (2009, 2015) call the “simple argument”, which have a normativist conclusion. Nevertheless, as we will see in §2.2.4, versions of problems identified by Glüer and Wikforss for the simple argument for normativism mar the content argument as well.

### §2.2.1 The extended argument

Many of the authors who make use of the content argument are normativists (Boghossian, 2003; Gibbard, 2003, 2005; Shah, 2003; Wedgwood, 2002). Wedgwood applies the content argument explicitly to the question of intrinsic normativity (Wedgwood, 2002):

P1) A belief is correct iff it is true

P2) Correctness is normative for belief

P3) It is an epistemic norm for belief that it is correct iff it is true

P4) Epistemic norms are essential to the types of states they apply to

C) It is essential to the state of belief that it is governed by the epistemic norm of correctness iff true.

For Wedgwood, normative concepts entail commitments and obligations:

If the concept ‘F’ is normative for a certain practice in this way, then engaging in the practice commits one to treating the judgment that a certain move is not F as representing a decisive reason against making that move (at least if there is an available alternative move that one judges to be F). To say that there is a “decisive reason” for one not to make a certain move is to say that one (in some sense) “ought not” to make that move. So, engaging in that practice commits one to accepting that one (in some sense) “ought not” to make moves within the practice that are not F. (p. 298)

[C]onsider the judgment ‘It would be incorrect for me to believe p’. This judgment commits one, not merely to trying to bring it about that one does not believe p, but directly to not believing p. (p. 270)

This is what it means for correctness to be a normative concept for belief: it is regulative of our practice of reasoning, and it commits believers to behave in certain ways. The next step in the argument is that the content of the correctness condition is an epistemic norm for belief. An epistemic norm, on Wedgwood's framework, is a necessary general principle that links a normative concept for epistemic practices to the non-normative circumstances under which those concepts apply, assuming that normative properties supervene strongly on truths that can be stated in entirely descriptive terms. For the normative property of correctness, this is the principle that links correct beliefs to true beliefs with a biconditional, i.e. the correctness condition (CC). Epistemic norms, according to Wedgwood, are necessarily true. The final step in the argument from correctness, is that epistemic norms (like CC) are essential to the mental states they range over:

There are reasons for thinking that these epistemic norms articulate essential features of the types of beliefs that they apply to ... First, as I have defined the term, "epistemic norms" are always necessary truths; so in that sense, these norms articulate necessary features of the types of beliefs that they apply to. Second, it seems plausible that, quite generally, types of mental state are individuated by the conditions under which they satisfy normative concepts. Suppose that there were two distinct types of mental state that did not differ in any way with respect to the conditions under which they satisfy normative concepts. Then these two types of mental state would count as correct under exactly the same conditions; they would count as rational under exactly the same conditions; they would be supported by exactly the same reasons, and would constitute reasons for exactly the same further actions and attitudes, under all possible circumstances; and so on. But in that case, it is very hard to see how these really could be distinct types of mental state at all. (p. 270)

Let's look at the second reason first. Mental states, Wedgwood writes, have normative application conditions, and these are unique to the type of mental state in question. That means that mental states are individuated by their norms, and hence we can expect them to be

essential to those mental states. Much will be said (in Chapter 4) about the question of state individuation through normative criteria. As I will argue there, it is far from clear that belief can be distinguished from other mental states through its norm, and it certainly cannot be distinguished via the correctness condition that Wedgwood argues is essential to belief as the most fundamental epistemic norm. For now, I want to briefly comment on the last step of the argument from norm uniqueness: that if a type of mental state can be individuated from other types of mental states through the norm that governs it, that norm is essential to the mental state in question. Suppose Wedgwood is right and its correctness condition individuates belief: anything the conditions of correctness of which are given by CC is a belief. Compare it with the following norm: "Marbles ought to be sorted into the box labelled "blue marbles" if and only if they are blue". It seems very plausible that this norm is true only of marbles: nothing else ought to be sorted into the box labelled "blue marbles" if and only if it is blue. But the sorting norm is not essential to the nature of marbles. Individuation through a norm does not seem sufficient in itself to establish intrinsicness. Presumably the normativist would want to say that the sorting norm is somehow relevantly different from the correctness condition, perhaps because only the latter is necessarily true. That brings us to the next section.

The other reason Wedgwood gives us for the belief-intrinsic nature of epistemic norms is their supposed necessity. It is not clear what Wedgwood has in mind precisely. The idea seems to be that if a principle is a necessary truth, the ascriptions made in it express essential features of the things that are being ascribed to. Take Wedgwood's principle "A belief is correct if and only if the proposition believed is true". This ascribes CC to belief, and according to Wedgwood CC is necessarily true, so it expresses an essential feature of belief. So far so good, but now look at the equivalent principle "A believed proposition is true if and only if the belief in it is correct". Is

the truth condition (true iff correct) an essential feature of believed propositions? This seems like a strange result. If propositions have essences, then that subset of them that is believed by someone won't have a different essence than those that no one is entertaining: propositions don't change their nature depending on whether they fall into the believed set or not. Whether a proposition is believed or not is contingent, but its essence is something it has by necessity. And it seems very unintuitive that it would be part of the essence of propositions *generally* that they are true if and only if they are true-making for beliefs. Presumably, propositions could exist in a possible world without beliefs, so propositions' essences shouldn't make reference to them.<sup>14</sup>

In addition to this specific problem, I mentioned in §1.3.2 that essences are unlikely to reduce to what is necessarily the case of the things they are the essence of. The normativist should therefore do some backing up of the link between the necessity of the norm of belief and its essential status. Wedgwood has later on (Wedgwood, 2009) defended normativism with a different and much more explicit argument, to which we turn in §2.4. But his idea we encountered here of the necessity of the epistemic norm is indicative of much of the literature: CC is tremendously popular as a candidate norm of belief, and since it is hailed as obviously true by those who defend it, postulating it as part of belief's nature seems to be for many an easy step to make. The following is conjecture, but the scarcity of positive arguments for doxastic normativism may well have something to do with the widespread success of CC, which

---

<sup>14</sup> On a formulation of CC "A belief that p is correct is and only if p is true", we get "A proposition p is true if and only if a belief in p is correct". On that formulation we can skip the part about the potential difference in essence between believed and non-believer propositions and go straight to the last part of the argument: the essence of propositions should not make reference to belief, for propositions could exist in a world without belief.

is often presented as impossible to argue with. If there were more mainstream disagreement about the content of the belief norm (the norm criterion), perhaps the idea that said norm is intrinsic to belief would strike philosophers as less obvious. In any case, we should spend at least a little time in this thesis investigating the content of the belief norm, and this is the right place to do it. CC is popular among normativists, and its supposed necessary truth has been used explicitly in a defense of doxastic normativism. Yet the correctness condition of belief has been seriously misconstrued.

### §2.2.2 Belief's correctness condition

Let us start with a brief look at the way CC has been endorsed by contemporary authors. Allan Gibbard (2005) writes: "For belief, correctness is truth. Correct belief is true belief. My belief that snow is white is correct just in case the belief is true, just in case snow is white. Correctness, now, seems normative." (p. 338) Gibbard's paper is meant to explain how this normativity arising from truth is possible without violating the is/ought distinction. He offers such an explanation by appealing to the difference between subjective and objective oughts, and showing how the latter can be characterised in terms of the former. But nowhere in his whole endeavour does Gibbard give us any reasons why we should think that truth is the normative correctness condition for belief in the first place. Both CC as well as its normative import are simply asserted in the beginning of the paper, and then taken as given throughout. As we will see momentarily, this is a common pattern.

A different kind of norm of belief is defended by Daniel Whiting (2010)<sup>15</sup>. He writes that “For any  $p$ : the belief that  $p$  is correct if and only if it is true that  $p$ .” (p. 213). Similarly to Gibbard, he offers no argument to this extent but he does mention that truth is the most “obvious candidate” for belief’s correctness condition. Whiting then states that CC has a “normative dimension”. Much of his paper is devoted to establishing what this normative dimension is. He directs his findings against the view (which we implicitly find in Gibbard, but also in many other places (Boghossian, 2008; Engel, 2013c; Wedgwood, 2007)) that correctness is a matter of obligation in the sense of “ $\phi$  is correct if and only if one ought to  $\phi$ ”. Correctness, Whiting argues, actually has little consequence for what one ought to do, and is much more adequately expressed in terms of permissibility “ $\phi$  is correct if and only if one may  $\phi$ ”. Consequently, he endorses the following norm of belief: “For any  $S$ ,  $p$ :  $S$  may believe  $p$  if and only if it is true that  $p$ ” (2010, p. 216).

A paper that follows a similar trajectory as Whiting’s but with a different conclusion is McHugh’s (2012). He writes that “The standard of correctness for a belief is the truth of the belief’s propositional content. ... The question I want to address is: in what sense might the standard of correctness for belief be normative?” (p. 9). The right sense of normativity according to McHugh is evaluative as opposed to deontic. McHugh formulates a truth norm of belief in evaluative terms: “For any  $S$ ,  $p$ : if  $S$  believes  $p$ , then that belief is a good doxastic attitude to have to  $p$  if  $p$ ”

---

<sup>15</sup> Daniel Whiting has expressed sympathy for the normativist project but has not (at least not in the works of his I quote) explicitly endorsed doxastic normativism. In his (2013) he writes: “Some suggest that it is constitutive of an attitude’s being one of belief that it is subject to a norm of truth. I shall assume that, if the norm holds, it does so necessarily but I shall not take a stand on what this might tell us about the nature of belief. That is a topic for another occasion.” (p. 123) I use Whiting here as someone who uses the content argument for the truth criterion of belief to illustrate the widespread acceptance of CC.

is true, and that belief is a bad doxastic attitude to have to  $p$  if  $p$  is false". *That* the standard of correctness is normative, and that it is marked by truth, is taken as given.<sup>16</sup>

### §2.2.3 Correctness and normativity

As we can see, significant debate exists about what kind of normativity (obligation, permission, positive evaluation) is involved in belief's standard of correctness. This happens against the backdrop of an explicit acceptance of the correctness condition for which no direct motivation or warrant is offered. Clearly, these authors think CC enjoys sufficient obviousness to be in no need of defense. Perhaps the relationship between correctness and truth is analytic. Take a sentence like "Darwin's belief that a moth exists that coevolved with the Christmas orchid turned out to be correct". Here, the correctness of the belief seems to entail its truth. Additionally, the correctness of Darwin's belief seems to require little more than that the belief turned out true (the sentence as stated seems to state nothing about Darwin's epistemic situation) so that we may think that it is correct *whenever* it is true. I take it that the authors who endorse CC would think that the way correctness functions in the sentence about Darwin's coevolution prediction is paradigmatic of the correctness of beliefs, i.e. that whenever a belief is correct it is true. Such observations about consistent patterns of understood implication of the words "correct belief" in ordinary language should indeed have a bearing on our philosophical theory of the concept of doxastic correctness. However, as will quickly emerge further down, this pattern is far from established. Additionally, even if it were the case that whenever one says "correct belief" one is

---

<sup>16</sup> In a later paper (2014) McHugh argues that the right way to characterise correctness is actually fittingness or appropriateness, which according to him is neither a prescriptive nor an evaluative norm. In that paper, he writes that he uses truth as the standard of correctness "for brevity's sake" and would actually like his argument to be compatible with knowledge as the standard of correctness as well. My objections below should therefore not be counted against that paper.

referring to a true belief, *CC qua normative principle* does not follow. For it is still a possibility that “correct” in this context simply means “true”. If the “implication” of truth in correctness predications to belief were due to synonymy, CC would be completely trivial and much work would be needed to show how anything normative could follow from it. From now on, when I speak of CC, I mean the correctness condition qua norm of belief (or founding principle of a norm of belief): that truth makes belief correct where “correct” is read as a normative property beliefs ought to have. I will argue first that patterns of correctness ascription to belief in ordinary language do not provide *evidence* for CC because they involve a use of the word “correct” that is synonymous with “true”. Secondly, I will argue that CC is false.

#### §2.2.4 The semantics of correctness ascription

Before proceeding to my arguments for this thesis about the semantics of correctness ascription, it will be useful to situate my view in reference to some authors who have addressed the meaning of the word “correct”. Defenders of the argument from correctness are aware of the worry that CC could be read as trivial due to synonymy. Wedgwood (2002) offered two reasons why “correct” should not be read as “true” in his formulation of the correctness condition. Vahid (2006) has convincingly argued for the failure of both of them. In more recent work (2007), Wedgwood has repeated that “correct” is not a synonym for “true”, since it would be “a solecism to describe a sentence or a proposition as correct” (p. 157). Bykvist and Hattiangadi (2007) were more influential in their complaints about CC than Vahid. They write that CC “looks trivial, particularly if we interpret ‘correct’ as synonymous with ‘true’ and take ‘belief’ to refer to the proposition believed.” (p. 277) Furthermore, they argue in detail that a

normative reading is not available, because it would lead to implausible demands by the resulting norm regarding blindspot propositions.<sup>17</sup>

I agree with the existing arguments (Bykvist & Hattiangadi 2013, Schroeder 2003, Vahid 2006) to the effect that it is not clear that “correct” cannot mean true. While Wedgwood claims that this is not a meaning of “correct”, Whiting (2010) seems to allow for the possibility that it is, but claims that “correct” *also* has a “normative dimension”, so that CC remains useful as long as it has at least one normative meaning. The problem for that strategy however, is that as soon as we allow that CC has a trivial reading, the case for the normative reading is seriously undermined, as I will explain below.

Some anti-normativist views of the semantics of “correct” have been proposed in the literature (Hattiangadi, 2006, Bykvist & Hattiangadi 2013, Glüer & Wikforss 2015), but my view differs from these in a number of ways. Bykvist and Hattiangadi (2013) agree with the defenders of the correctness arguments that CC is true: “the appropriateness of applying ‘correct’ covaries with the appropriateness of applying ‘true’” (p. 104). They deny however that CC is a normative principle. “Correct”, according to Bykvist and Hattiangadi, means something like “in accordance with a contextually salient standard x”. In the context of evaluating beliefs, that standard is truth, which is why ascribability of truth and correctness covaries. Since, according to them, there are scenarios where we call something “correct” without affording it genuine positive evaluation, correctness ascriptions are not *genuinely* normative. Correctness is not a normative concept on their view, and hence nothing normative follows from CC. Hattiangadi had already argued in

---

<sup>17</sup> They spell this out in terms of different ways to formulate an ought reading of CC, and later (2013) supplement this with blindspot arguments against alternative normative analyses of correctness. The blindspot objection against CC is a forceful one, but my own arguments against CC are independent of it.

earlier work (2006) that correctness is not normative. But on her view, similarly to Glüer and Wikforss' (2015), only prescriptive normativity counts, whereas I address evaluative norms as well. So one difference between the view I will propose and that of authors like Bykvist and Hattiangadi or Glüer and Wikforss is that a candidate-meaning of "correct" like "in accordance with a standard" would count as a normative meaning. Another differentiating marker of my view is that I deny that applicability of "correct" and "true" covaries when talking about belief. In the rest of this section, I will argue that the word "correct" has no single meaning, not even a context-dependent one. Instead of context-dependent, I argue that "correct" is an ambiguous term with at least two meanings. The first is simply "true", not because truth is some contextually salient standard, but categorically. Under that meaning, "correct" can be applied to a subset of propositional attitudes, to assertions, or to answers. The second meaning of "correct" is something along the lines of what Bykvist and Hattiangadi suggest: "in accordance with a standard, appropriate". Under that meaning, correctness can be predicated of many more entities, including many forms of behaviour and the results of actions.

Dictionaries list (at least) two different meanings for "correct": "proper or appropriate in a given situation" and "true or accurate". If these constitute truly separate meanings, "correct" would be an ambiguous word.<sup>18</sup> Some evidence can be found through ambiguity tests for an interpretation of "correct" as semantically ambiguous. One such test is to combine two sentences that both feature the word in question in contexts where different meanings are encouraged, but use that word only once in the conjoined sentence. An example of such a

---

<sup>18</sup> In both the Merriam Webster and the Oxford English Dictionary, the two different meanings are put under a single lexical entry. In the Merriam Webster dictionary, this treatment is given to known ambiguous words like "date" as well. In the OED, these clear cases of ambiguity get multiple lexical entries. Polysemes like "man" or "wood" are treated under one lexical entry, the same way "correct" is.

conjunction reduction is “The colours and the feathers are light”. If “light” is ambiguous, the result of the conjunction reduction should sound infelicitous, which indeed it does. Conjunction reduction seems to have similar results for “correct”. Take first an example of conjunction reduction where “correct” is applied in the same sense twice: “In order to be qualified as a letter by the Royal Mail, the envelope needs to have the correct size and weight”. I take it this conjunction reduction is unproblematic. Take next a case where “correct” is applied under the purported different meanings. To make for a setting where it is plausible to speak of correctness in these two different senses, we need to provide a little more context than for the envelope: “Many of the people who applied had formatted their resumes incorrectly. Another problem was that the majority of candidates believed incorrectly that the internship was paid. The person we eventually hired had the correct resume format and belief”. The last sentence, which uses conjunction reduction, sounds infelicitous. This can be interpreted as a symptom of ambiguity of the term “correct”.

Another test for ambiguity stems from sentences where two ideas are opposed, using the same word. If it is possible to read the result as non-contradictory, that is evidence for the ambiguity of the word in question. Take, for example, “It is a dog but it is not a dog”, a textbook example of an ambiguity test.<sup>19</sup> At first glance, this sounds contradictory. There is however one interpretation on which it is not<sup>20</sup>, and we can bring it out by providing some context: “In the English language, the word “dog” can refer to any member of the canine species, but it can also refer to male members of that species only. Pointing to a golden retriever bitch, you could

---

<sup>19</sup> Zwicky and Sadock (1975)

<sup>20</sup> That is of course not to say that it does not violate some conversational principles.

therefore say "That is a dog, but it is not a dog". If "correct" is ambiguous, similar usage should be possible there. I think the following parallel case sounds at least not worse than the dog example: "In the English language, the word "correct" can mean "true", but it can also mean "appropriate". Reprimanding someone for their rudeness, you could say "Your statement that she looked tired was correct, but it was not correct".

All of this makes it highly likely that "correct" is an ambiguous term, and that its alethic dimension and its normative dimension correspond to different meanings of the word. I think this is plausible already from the ordinary usage of the word and the two ambiguity tests above should go some way to demonstrate it. If "correct" is indeed ambiguous in this way, that makes the content argument from correctness inconclusive. Recall the content argument:

P1) A belief is correct iff it is true

P2) Correctness is normative for belief

C) It is an epistemic norm for belief that it is correct iff it is true

If "correct" is ambiguous between "true" and "appropriate", then P1 is only established on the first meaning, on which it is trivially true. This provides no evidence for a reading of P1 on which "correct" has the normative meaning (i.e. CC). P2 might be true on the second meaning of "correct, but on that meaning P1 stands without argument. Therefore, C does not follow from the truth of P1 and P2, for the meaning of the term "correct" is not held fixed throughout. We need a different argument for CC (qua normative principle). But such an argument might be difficult to find, because once we make the normative meaning of correctness explicit, there are counterexamples to CC. Below are three examples of contexts in which "correct" or "incorrect" are used in a way that is at odds with CC:

Ex. 2.1: Hans has a number of beliefs that he formed in various incorrect ways. He believes he's a great singer because of wishful thinking, he believes he's going to die in a car crash because his medium told him that, he believes the answer to some multiple choice question is the fourth option because 4 is his favourite number. Of these incorrect beliefs, two are false.

Ex. 2.2: Irene struggles to recover from her alcoholism. Her counsellor thinks that part of the problem is Irene's pessimistic idea that it "doesn't make a difference" whether she is an alcoholic or not. The counsellor compares Irene to another recovering alcoholic: Joe. The counsellor says: "Joe has all the correct beliefs to aid him in his recovery. He thinks the world is a better place if he doesn't drink, that his sobriety makes up for his past sins and that his steadfastness will grant him forgiveness. All that is really beneficial for Joe, whereas Irene's attitude, although more realistic, is counterproductive."

Ex. 2.3: For many children, evolutionary theory takes time to get used to. At this school, we prefer to teach this subject by phases that build on one another as opposed to introducing all the concepts at once. In last year's group there was a case that illustrates the importance of this approach quite well. One girl had read about genetic drift somewhere outside of school and consequently started believing that chance is an important factor in evolution. Although that is strictly speaking true, it ended up confusing her to the point where she didn't understand the concept of natural selection anymore. At that point in the curriculum, the correct thing for her to believe would have been that evolution happens exclusively through adaptations.

In Ex. 2.1, it is asserted that Hans has a belief that is incorrect yet nevertheless true. In Ex. 2.2, Joe's beliefs are labelled "correct" even though the speaker distances herself from their content and implies that she takes them to be false. In Ex. 2.3, a belief that is acknowledged to be true is called incorrect. In all three cases, whatever criterion the speaker seems to have in mind for the correctness of the belief under evaluation, differs from truth. In the first case, by "incorrect" the speaker seems to refer to an epistemic deficit. In the second case, correctness is marked by personal beneficence through the beliefs' aid in recovery. In the third case, what is correct is understood in the context of fitting into specific stages of a learning process. If truth is the only thing that can determine correctness qua normative concept for beliefs, these uses of the word should all be completely unacceptable. Yet they seem perfectly intuitive in the contexts as presented. Of course, examples like 2.1-3 are rare. In most cases where we speak of beliefs as correct in the normative sense, those beliefs are true. Yet there is an explanation for this, and it is not in the spirit of CC.

#### §2.2.5 More ambiguity

"Belief" is ambiguous between the state of believing and the proposition that is believed. Some predicates disambiguate "belief", at least to an extent. Calling a belief stupid, for instance, clearly refers to the state of believing, as propositions cannot be said to be stupid. Calling a belief incomprehensible makes it clear that one finds it incomprehensible that somebody would hold that belief, not that one takes the proposition involved to be incomprehensible (as long as everybody involved accepts that incomprehensible things cannot be believed). When philosophers talk of beliefs as occurrent or dispositional, they also have the state of believing in mind. On the other hand, some predicates seem to refer quite clearly to the proposition

believed. If I say “The patient holds many trivial beliefs”, what I mean is that the patient believes many things that are trivial, not that there is something trivial about the mental states involved. Indeed, discovering someone who actively holds before her mind a host of trivialities may be the very opposite of trivial from a psychological point of view. Similarly, if I call someone’s beliefs revolutionary, I am probably talking about the content of their beliefs, not about the fact that they hold them (take a subject who has been born and raised in a community of revolutionaries; there is nothing revolutionary about her having those mental states).

With these distinctions in mind, return to CC. The correctness condition is sometimes defined in terms of “believing that p” (McHugh & Whiting, 2014) which clearly refers to the state of believing and sometimes in terms of “the belief that p” (Gibbard, 2005; Whiting, 2010) which is ambiguous between state and content. CC is clearly concerned with the state of believing though, as it serves to evaluate people’s mental life normatively. In most of the literature on the norms of belief, “belief” is to be understood in this sense. However, and this has problematically been overlooked by the defenders of the content argument, “correct belief” is most naturally read as talking about the content of beliefs. If we want to make it clear that we are evaluating the belief state, we can make that clear by saying “correctly believing p” or “it is correct to believe p”. If we simply talk about a “correct belief” we risk that being understood as referring to a correct proposition being believed, not to the correct belief to have.

Truth is the main aspect we evaluate propositions for. Propositions are furthermore unlikely to have normative properties<sup>21</sup>. Therefore, when we say “the belief that p is correct”, if “belief” is read in the content sense, that will reinforce the disposition to read “correct” in the alethic sense.

---

<sup>21</sup> Except on views where meaning itself is normative, see §3.3.

In line with that, if we use the formulation that make it clear we are evaluating the correctness *of the mental state*, counterexamples to CC are much easier to find.

Ex. 2.4: You were incorrect to believe he was guilty before you had any evidence.

Belief was not the correct attitude at that stage of the inquiry.

It is incorrect of you to believe the elevator is safe to use in a fire, even if that should turn out to be true.

It was correct of you to believe the animal was a zebra, even though it was a painted mule.

If (the plausibility of) the case for CC depends on a reading of "correct belief" as an evaluation of the content of the belief, that makes CC hardly of use to the endeavour of grounding a norm for belief states.

### §2.3 The dilemma of correctness conditions

For the above reasons, I am highly sceptical of the principle that beliefs are correct (in a non-trivial, normative sense) if and only if they are true. But of course, I have little company in this scepticism. Almost all the normativists (narrowly understood) that I am attacking accept CC qua normative principle. But CC would make for problems for normativism even if it were true. If it is the case that beliefs are correct whenever they are true, and if it is true that the correctness at hand is normative, and if the norm of belief follows from those two facts, there is a counterintuitive consequence that is a serious problem for doxastic normativism. This consequence has to do with the normativity of weak affirmative attitudes.

Recall that affirmative attitudes are propositional states with a mind-to-world direction of fit that involve an endorsement or affirmation of the proposition by the subject. Affirmative attitudes

include believing, guessing, suspecting, conjecturing, estimating, and assuming. Weak affirmative attitudes are the affirmative attitudes other than belief. All the cited examples of affirmative attitudes can be said to be correct or incorrect. And it would seem that in so far as truth makes beliefs correct, it is the thing that makes these other attitudes correct as well:

Ex. 2.5: You guessed correctly.

It turned out my suspicion was correct.

There is a prize for the correct estimate.

Of course, for reasons given above, I don't think truth is the standard of correctness of belief, indeed I think it unlikely that belief has a single standard of correctness. Truth is clearly *a* standard: there is a reading of "beliefs are correct iff true" where "correct" is not interpreted normatively. That statement would be trivially true, but does not represent a *standard* for belief, merely a truth condition. On the normative reading, there are cases where truth makes belief correct, but there are others where some other thing, like respect, or politeness, or evidence, makes it correct. Truth is thus a standard of correctness, but not a privileged one. The same is likely to be the case for the weak affirmative attitudes: exceptions to the truth standard can be found for them too.

Ex. 2.6: I don't care if it's true, it's not correct for you to suspect your superior of fraud with the kind of evidence you have.

You were lucky that the gun was indeed unloaded, but guessing that was not the correct thing to do.

This indicates that correctness conditions function very similarly for the weak affirmative attitudes: usually, we call them correct when they are true, but in some contexts the standard

of correctness is a different one. The defender of the content argument will want to argue that the exceptions to the truth standard are somehow misleading. Whichever strategy she uses to that effect will apply to Ex. 2.6 as well, so she will have to accept that the correctness condition applies to the other affirmative attitudes to the same extent as it does to belief.

It is sometimes said that the correctness condition is *unique* to belief, so that it individuates belief from other mental states (Shah & Velleman, 2005; Wedgwood, 2002, 2007a). This is a strange suggestion, given how natural the locutions in Ex. 2.5 are. Granted, Shah and Velleman don't consider the attitudes in those examples, but use imagining as their example instead (which is not an affirmative attitude).<sup>22</sup> For the real affirmative states, their aptness for correctness or incorrectness seems undeniable, as seems the extent to which it is truth that makes them correct. "Correct guess" has about 193.000 google hits, "correct assumption" has about 143.000, whereas "correct belief" has about 77.800. "Suspected correctly" still gets about 28.100. So the normativist will need a sophisticated argument for why we should think that weak affirmative attitudes do not have truth as their correctness condition the way belief does. This certainly cannot be simply assumed.

This, then is the *dilemma of correctness conditions*: to make the argument from correctness work, one has to either 1) argue that belief has different correctness conditions than other affirmative attitudes, or 2) endorse the same epistemic norm for all affirmative attitudes. The problem with 1) is, that even if such an argument can be found, it would flatten the entire

---

<sup>22</sup> Speaking of imagination as "correct" or "incorrect" is certainly a lot less natural than for the affirmative attitudes. But even for imagining it's not impossible to construct a context where we can: "I asked him to imagine what it would be like to be a single mother, and he imagined it incorrectly as similar to being single but with extra household help".

intuitive appeal of the correctness condition for belief, which rests in natural language practices of ascriptions of correctness to beliefs. As we have seen, there has been no intuition- or linguistic-practice-independent argument for the premiss that beliefs are correct if and only if they are true. In order to avoid 1), the defender of the argument from correctness will have to accept 2). For if belief's correctness condition makes for the epistemic norm for belief, we should expect the correctness condition of the other affirmative attitudes to also generate epistemic norms for them. And since they have the same correctness condition, we should expect them to be subject to the same epistemic norm. This seems wildly implausible. Guesses, assumptions, estimations etc seem like the kinds of mental states whose purpose it is to be adopted in exactly those situations where belief is not normatively available.<sup>23</sup> Secondly, if all affirmative attitudes had the same normative application conditions, we could never criticise someone for their *choice* of affirmative attitude.<sup>24</sup> But it seems that we can:

Ex. 2.6: You shouldn't have guessed, you should have estimated.

Conjecturing that global warming is real is not enough, you should believe it.

It's ok to have suspicions, but you should not believe the defendant is guilty until after the verdict.

In conclusion, establishing state normativism via the argument from correctness is not a promising avenue for the normativist. Deriving essential truths about the nature of belief from the necessary truth of the norms that govern beliefs is metaphysically tricky (see §1.3.2). Furthermore, if that norm is supposed to be either the correctness condition, or a normative

---

<sup>23</sup> Further reasons why a shared correctness condition for all affirmative attitudes is highly problematic are discussed in §4.3.

<sup>24</sup> I will argue in Chapter 7 that this kind of normativity that recommends certain affirmative attitudes over others is an Achilles heel for doxastic normativism generally.

translation of the correctness condition (like the ones the content argument provides), serious difficulties loom for the favoured criterion of correctness, from semantic considerations as well as from counterexamples. Finally, on any given criterion for the correctness condition, the dilemma of weak affirmative attitudes makes for deeply unattractive consequences of the argument from correctness.

## § 2.4 The argument from dispositions

Recall the quote from Kate Nolfi in §2.1.2, which states that there are “surprisingly few sustained defences of normativism about the nature of belief”. Nolfi herself mentions a few (Shah, 2003; Shah & Velleman, 2005; Wedgwood, 2007a; Zangwill, 1998, 2005, 2010). Nishi Shah and David Velleman are defenders of conceptual normativism and I discuss their argument in detail in Chapter 4. That leaves Wedgwood, Zangwill and Nolfi herself. They all discuss versions of what I call “the argument from dispositions”. I will present an objection to this argument, using Wedgwood’s formulation<sup>25</sup>.

The argument from dispositions starts with the idea that types of attitudes are differentiated by the dispositions that give rise to the adoption and abandonment of those attitudes:

What makes it the case that an attitude-type in your repertoire is the attitude of intention? According to this sort of dispositionalism, it is in virtue of some of your dispositions with respect to an attitude-type in your repertoire that this attitude-type

---

<sup>25</sup> Wedgwood actually argues for much more than normativism about belief in the two papers I reference. He seeks to establish normativism about the intentional generally. His strategy is to argue for normativism about concept possession: in order to possess a certain concept, one must be disposed to apply it in accordance with certain norms. From there, he proceeds to argue that the intentional attitudes towards those concepts are normatively constituted as well. When speaking of “Wedgwood’s argument”, I mean the argument for the normativity of the intentional as applied to belief specifically (which Wedgwood does mention himself explicitly as well).

counts as the attitude of intention; and this disposition is essential to being capable of the attitude of intention. (Wedgwood, 2007, p. 90)

For belief, the relevant dispositions will be things like the disposition to form the belief that there is a certain object in front of one when presented with visual evidence of that object under good perception conditions, believing the content of reliable testimony, abandoning a belief when discovering one's reasons for holding it were faulty, abandoning at least one of two beliefs that are discovered to be incompatible etc. Let us for present purposes grant the premiss that beliefs and other propositional attitudes are marked by dispositions.<sup>26</sup>

#### §2.4.1 Essential dispositions

The next premiss expounds on the dispositions that are essential to the capacity for belief. These dispositions, Wedgwood writes, must be rational dispositions. To show why, he asks us to imagine a perfectly rational being who has no irrational dispositions of any kind. This rational being, he writes, should be assumed to possess the same concepts I possess, and to share my capacity for my intentional attitudes towards those concepts. Otherwise, the rational being would not be able "to ascribe attitudes to [me] accurately, and to diagnose the various confusions and irrationalities that mar [my] thinking" (p. 90). For the perfectly rational being, her possession of those concepts and attitudes must be a consequence of rational dispositions, for she has no irrational ones.

The third premiss in the argument from dispositions is that rational dispositions must be characterised in normative terms. The reason for this, per Wedgwood, is that rational reasoning needs to be defeater-sensitive:

---

<sup>26</sup> We will encounter serious objections to this assumption in Chapter 6.

So the only way to specify a form of reasoning that will always count as rational is for one's specification to include a proviso that requires the absence of such defeating conditions...But the very notion of a "defeating condition" is a normative notion. (p. 91)

Wedgwood thinks we cannot define rational dispositions in entirely descriptive terms, because "forms of reasoning" are only rational if they take defeaters into account, and that clause must be specified in normative terms.

To sum up Wedgwood's argument:

- P1) Having beliefs is a matter of having certain dispositions.
- P2) The dispositions essential to belief are rational dispositions.
- P3) Rational dispositions are normative.
- C) Beliefs are normative.

#### §2.4.2 Resisting the normativity of rational dispositions

What should the anti-normativist say in response to the argument from dispositions? I will argue that there are reasons to reject both P2 and P3. Additionally, I will argue that even if the normativist can escape these objections, the two premises in conjunction have consequences she should avoid, and that she therefore should give up at least one of the two. Let us start by looking at P3: rational dispositions are normative.

The reason to think that rational dispositions must be characterised in normative terms is that they should involve sensitivity to defeaters. The normativists contend that defeaters must be defined in normative terms, and that the conditions on rational dispositions that are meant to protect belief against defeaters can therefore also only be defined in normative terms. According to Wedgwood, any form of reasoning that does not include normative elements will be in principle defeasible; he gives the example of the form of reasoning "accepting inferences

form modus ponens”, which he writes can still be defeasible, for instance, by a large number of experts testifying against the inference one draws.

Even if we accept the assumption that defeaters can only be described normatively<sup>27</sup>, I don't think P3 follows as easily as the normativist makes it look. The problem is that Wedgwood reasons from the need for “forms of reasoning” to be defeater-sensitive, to dispositions to believe needing to be defeater-sensitive. This step is problematic. Why does it have to be the individual dispositions to form or revise beliefs that need all be sensitive to defeaters, instead of the “form of reasoning” as a whole, which consists of multiple dispositions? On the latter picture, a perfectly rational being might have one disposition to form a belief that p whenever she sees p in front of her, and another disposition to suspend judgment about p if she knows there to be reasons not to trust her eyes in the situation at hand. The result will be defeater-sensitive belief formation. Rationality then, would not be a property of individual propositions, but a property of their interactions.

I think there is good reason to prefer this picture to one where the *ceteris paribus* conditions are packed into the individual dispositions. Firstly, from introspection in cases of forming or abandoning beliefs in reaction to defeaters, it seems that this is a much more natural account of belief formation and abandonment in reaction to defeaters. Take for illustration the scenario where I see an object before me that appears red, but know that the object is illuminated by a red light. As a result, I suspend belief about whether the object before me is red. On Wedgwood's account of dispositions, I have a disposition to form beliefs that there is an XYZ in front of me when presented with the visual input of an XYZ in front of me, except if there are

---

<sup>27</sup> See McHugh & Whiting (2014) for discussion.

defeaters present to my visual evidence of XYZ. In this scenario, that disposition to believe is then not activated. Instead, some other disposition I presumably have is activated, resulting in suspension. However, it seems to me that in cases like these I do *experience* a disposition to believe that there is a red object before me. The fact that I end up not adopting the belief in question is due to the manifestation of the disposition to suspend belief in reaction to my knowledge of the red spotlight: the two dispositions have incompatible outcomes, and in this case the correct one wins. The same phenomenon seems to occur in a case where I perceive a possible inference from modus ponens but simultaneously know of expert disagreement with the inference. On Wedgwood's picture, my disposition to believe inferences from modus ponens would just not be activated at all in this case, because of its anti-defeater condition. Yet we clearly feel the pull of the modus ponens in such situations: it feels like we are *resisting an inclination* to believe the inference, not like we are simply not inclined to believe it because of the defeater.

So we don't have to have to accept P3, since we can account for the defeater-sensitive rationality of forms of reasoning without the appeal to defeater-clauses in the dispositions to believe themselves, an account that has the advantage of phenomenological plausibility. Yet even by the normativist's own light, there are difficulties with P3. If we combine Wedgwood's definition of a rational disposition (as one that takes defeaters into account) with the second premiss of his argument (that only creatures with rational dispositions are believers), P2 starts to look significantly less plausible. To see why, consider the following case:

Ex. 2.7: Defins is a creature that has all kinds of dispositions to form, maintain and revise attitudes in keeping with basic evidence and rules of reasoning, yet lacks any disposition

to pay attention to defeaters to her reasons to accept some proposition. Perhaps the creature in question is simply not equipped with the cognitive complexity necessary to reason about evidence. She forms endorsing propositional attitudes about the presence of red objects whenever she sees objects that look red, even when she is in a room filled with red light, although she abandons her endorsement when she sees the object look blue when picking it up. She forms endorsing propositional attitudes in response to testimony even when the testifier is known to her to sometimes lie, although not when the testifier is known to her to always lie. In other words, she is sensitive to evidence that rebuts her reasons to accept some proposition, but not to evidence that undercuts her reasons to accept it.

On the conjunction of P2 and P3, Defins is not a believer: if it is true that all rational dispositions must be defeater-sensitive, and one only has the capacity for belief when one has the rational dispositions essential to belief, Defins has no beliefs. There is something implausible to this conclusion. We have ended up with a very demanding definition of what counts as belief, although the normativists seemed to have started out talking about a very ordinary conception of belief. Keeping the implausibility of Defins' characterisation as incapable of belief in mind, let us move on to evaluating P2 of the argument from dispositions: that the dispositions essential to the capacity for belief are rational dispositions.

#### §2.4.3 Resisting the rationality of essential dispositions

Recall Wedgwood's reasoning for P2: someone without any irrational dispositions can share the attitude of belief with me.

[A]ny concepts that you have could be shared by a perfectly rational being who had no irrational dispositions at all...However, if it is in virtue of some of her dispositions that the perfectly rational being possesses these concepts, and the perfectly rational being has no irrational dispositions at all, then it must be in virtue of some rational disposition that she possesses the concept. (p 90)

The idea that someone without any irrational dispositions can share the attitude of belief with me seems highly plausible. Perfectly rational beings seem like ideal candidates for knowledge, and knowledge entails belief. The idea that something with only irrational dispositions does not count as believing also seems plausible. Imagine a system that has the two atomic propositional attitudes  $p$  and  $q$  as its states and manipulates them in accordance with two rules: " $p \wedge q \rightarrow \neg p$ " and " $(p \rightarrow q) \wedge p \rightarrow \neg q$ ". I take it that many will grant that the resulting cycle of updates of the truth value of  $p$  and  $q$  do not count as beliefs. Yet contrary to the way dispositional argument puts things, that in itself does not suffice to establish P2. The normativist needs a background assumption here: that every disposition to form or revise propositional attitudes is either rational or irrational. Call this assumption "binary rationality". Without binary rationality, the point that I can share the capacity for belief with a perfectly rational being only entails that the capacity for belief cannot be due to any irrational dispositions, leaving open the possibility that the dispositions essential to belief are arational.

But we don't have to accept binary rationality, especially not without argument. Many of our mental dispositions might be rationally neutral. If we adopt a view of rationality as a property of the interaction between dispositions, we can easily reject binary rationality about mental dispositions. Additionally, because rationality seems to have a lot to do with the relationships between different beliefs, it might not apply to a very simple belief system. A creature that is only capable of one propositional attitude at a time might not have any irrational dispositions,

but it is not disposed towards rationality either. If these singular attitudes are phenomenologically like beliefs and have the relationship to behaviour that beliefs do, I suggest they should count as beliefs, despite not figuring in relationships that would make them rational or irrational. If this is correct, then we have a counterexample to P2.

#### §2.4.4 Irrationality of defeater-insensitivity

Let us return now to the case of Defins in Ex. 2.7. On the conjunction of P2 and P3, Defins' dispositions to quasi-believe (which are identical to our dispositions to believe except for the fact that they do not take defeasibility of evidence into account) do not count as beliefs. This is because those dispositions to quasi-believe do not count as rational on Wedgwood's picture. But by binary rationality, if they are not rational, they are irrational. As a consequence, Defins has only irrational dispositions. That seems like an even worse result than Defins lacking the ability to believe.<sup>28</sup> It puts her into the same normative category as the system that only has contradictory dispositions like " $p \wedge q \rightarrow \neg p$ " and " $(p \rightarrow q) \wedge p \rightarrow \neg q$ ". Something has to give: if it is true that everything that is not rational is irrational, the criteria of rationality will have to be much more lenient than on the normativist's picture. If rationality really has to include defeater-insensitivity, it cannot be the case that everything that is not rational is irrational, and binary rationality fails. Without binary rationality, P2 is no longer established.

---

<sup>28</sup> It seems strange to call systems without beliefs irrational, for rationality and irrationality are normative evaluations of beliefs or of sets of beliefs. This is a consequence of normativist definitions of belief I will not be able to evaluate.

## §2.5 Resisting state normativism

In this chapter, we encountered some specific defences of state doxastic normativism. I focused on two: the argument from correctness and the argument from dispositions. Discussion of the former involved a detailed rebuttal of the content of the belief norm that has been favoured by most normativists: that true beliefs are correct beliefs (CC). I argued that CC is false: it is not the case that beliefs are always correct whenever they are true. I explained the reasons for CC's apparent, but misleading, plausibility. The falsity of CC does not disprove doxastic normativism directly, but it does weigh against its plausibility. We will continue to encounter views that appeal to CC, and obviously the arguments I presented in §2.2 will weigh against all of them: false normative principles are not essential to belief. I also argued that using *any* kind of correctness condition to define belief, even one that takes the data I listed which points to exceptions to the truth criterion into account, suffers from what I called the dilemma of correctness conditions: the normativist who wants to use correctness conditions to define belief will have to argue either that despite clear intuitions and linguistic data, belief has a different correctness condition than the weak affirmative states, or she will have to accept that all affirmative attitudes are normatively indistinguishable. In addition to these content worries, we noted that the demonstration of intrinsicness from necessary truth of a normative principle is not obvious, so even if CC could be replaced by another norm that does hold necessarily for beliefs, state normativism might not yet follow.

In the last part of the chapter, we evaluated the argument from dispositions, and I concluded that we need not be moved by it. We can reject P2 by denying binary rationality. We can reject P3 by arguing that defeater-sensitivity is not a property of individual dispositions to form beliefs.

Even if the normativist can provide supplementary arguments for binary rationality and against the system-view of defeater sensitivity, I have argued that she should give up at least one of P2 and 3, because of the implausibility of characterising a defeater-insensitive quasi-believer as entirely irrational.



## Chapter 3: Against State Normativism

### Non-agential beliefs, overdiscriminate definitions, and lawless deviance

As we have seen, the argument for state normativism from dispositions employs a premiss according to which the kind of rationality that is necessary for a creature to count as a believer involves sensitivity to defeaters. I argued that this requires a level of mental sophistication of the candidate-believer that seems too demanding: when imagining a creature who forms mental states based solely on input from confirming and rebutting evidence without paying heed to evidence *about her evidence*, conceiving of her as a believer is not counterintuitive. Indeed, it is plausible that there are some non-human animals that fail to be responsive to defeaters in this way. This points to a general source of potential problems for state normativism: the normativist wants to ground epistemic normativity in the nature of belief, but many of our epistemic norms are quite demanding of believers and require a certain level of sophistication to follow. Many candidate-believers' mental lives fall short of this kind of sophistication.

#### §3.1 Non-agential mental states

I use the term "non-agential candidate-believer" here as a denominator for non-human animals, some humans including young children, and machines. Perhaps calling these candidate-believers "non-agents" is problematic, but I will not go into detail about the nature of agency. "Non-agential" for my purposes is just a label for a set of beings, whether any or all

of them indeed fail to be agents, and if so which facts about them makes this the case should not be of much relevance to my argument. That most non-human animals, young children, and machines are not *moral* agents seems to be widely accepted in folk thought. Perhaps however what makes one an agent in the sense relevant to epistemology and philosophy of mind is different from what makes one an agent in the moral sense. If non-human animals, children, and machines are agents in a relevant sense, everything I say in this section but the label with which I encompass them should still apply.

### §3.1.1 Problems for non-agential attitudes

Non-agential mental states pose two kinds of problems for doxastic normativism. The first I call the *sophistication problem*. Because of a difference in cognitive sophistication, non-agents might conform to fewer epistemic norms than we do: there are things we expect of rational humans that many animals, for example, are not capable of. The sophistication problem mars arguments like the argument from dispositions that draw on requirements of rationality we recognise for ourselves (such as sensitivity to defeaters). Non-agential beliefs will pose a counterexample to a definition of belief as the type of mental state that is produced by dispositions that meet this kind of rationality requirement. A rabbit's belief might be a manifestation of dispositions to form a belief in response to olfactory evidence for the fox being on the left and dispositions to abandon that belief in response to evidence that contradicts the fox being on the left (say olfactory evidence that she's on the right or visual evidence that she's not on the left). But the rabbit might have no dispositions to respond to evidence that olfactory evidence might be unreliable in her present situation (say being very tired and smelling less accurately as a consequence). A rabbit may just believe "fox" whenever she smells fox, until

perhaps she has been wrong so many times that she becomes conditioned to ignore the smell. But learning to cease being sensitive to a certain kind of evidence is not the same as being sensitive to defeaters.

Of course, evaluated separately from Wedgwood's and Nolfi's disposition argument, normativism does not need to include reference to defeaters in its account of the nature of belief. The idea is usually that there is one norm that is intrinsic to belief, like the truth norm, from which other norms can be derived (like the evidential norm). Let's assume with the normativist that the strategy to ground the general epistemic normativity of belief in the truth norm works. Now you and I ought to take defeaters into account in our reasoning. If I go ahead and believe the present is hidden under the bed because I see a present-shaped shadow under the bed, knowing full-well that I'm very bad at recognising objects in dark conditions and that you tend to store all kinds of things under the bed, I'm being not entirely rational. I form a belief based on evidence I shouldn't have trusted quite so easily. On the normativist story, whichever norm it is that is intrinsic to belief explains the rationality norm that censures my hasty belief formation.

This is what causes the sophistication problem: if the intrinsic belief norm generates an obligation for me to take defeaters into account, by hypothesis it generates that same obligation in all believers, including the rabbit above. But the rabbit is not capable of that kind of sensitivity. Ought-implies-can intuitions might therefore mean she cannot be subject to that kind of norm. Even without a commitment to ought-implies-can, it seems clear that our intuitions about epistemic norms take feasibility constraints into account. To take an example, the reason why the norm of belief authors in this debate are looking for (with human epistemic agents in mind)

is thought to be dischargeable on fallible justification, is presumably feasibility-related. If what the norm obliges humans to is at least somewhat informed by feasibility, the same should be the case for animals. There is thus an explanation due by the normativist how the norm intrinsic to belief generates certain obligations in us, but not in rabbits. Such an explanation may well be available, but it should be made explicit if the normativist wants to make it clear that her account of belief is compatible with the existence of non-agential belief.

The sophistication problem has to do with norms we do not intuit non-agents to be subject to. A related problem focusses on norms of which it does seem plausible that non-agents are subject to them. Consider the following normative statements

Ex. 3.1:           The dog was wrong when he believed his owner beat him because he hadn't saved the lamb; his owner would have beaten him anyway.

The dog was mistaken when he believed his owner beat him because he hadn't saved the lamb; his owner would have beaten him anyway.

The cat believes mistakenly her litter is well-hidden, it is completely visible from above.

The bird erroneously believed the poisonous berry was edible.

Three different false beliefs are ascribed to animals, but more is said of their beliefs than that they were false: the animals in question are said to be mistaken or make errors as well, which are normative judgments. But they are quite thinly normative, and the evaluations in question seem appropriate to make of animals. Now compare them with these normative statements:

Ex. 3.2: The dog was wrong to believe his owner beat him because he hadn't saved the lamb.

The dog should not have believed his owner beat him because he hadn't saved the lamb.

The cat is wrong to believe her litter is well-hidden.

The bird ought not to believe that the berry is edible.

These statements seem much more problematic, and the problem stems from the kind of normativity through which the animals' beliefs are evaluated: these are ascriptions of violations of deontic norms. Even though animals can conform to, and in epistemically conducive situations tend to conform to, a truth norm, censuring them in the above way seems inappropriate. However, the same kind of normativity seems perfectly fine to ascribe to humans:

Ex. 3.3: William was wrong to believe opium would cure his illness; it made no difference whatsoever.

Matilda should not have believed her boss would keep his word; he ended up completely defrauding her.

You're wrong to believe you did the right thing; you caused great harm.

The spy ought not to believe the phone line is secure; it's tapped.

The same normative evaluations are made of humans as of animals in Ex. 3.2, and on plausible interpretations of the background stories for the three examples so that the reasons for the mistakes in Ex. 3.3 (rationalisation and wishful thinking biases, carelessness, and bad luck) are similar to the reasons for the mistakes in Ex. 3.2.

This phenomenon of a contrast in deontic ascribability poses difficulty for state doxastic normativism. I call this the *censure problem*: non-agents seem to be subject to less censure to epistemic norms than agents, even when these are norms they have sufficient sophistication to conform to. Now normativism is supposed to explain the epistemic normativity that governs epistemic agents, and that includes the normativity at work in Ex. 3.3. Our norms of belief should

include resources to censure false beliefs in this way, to say of epistemic agents that they were wrong to have certain beliefs or should not have had them. But as with the sophistication problem, if it is something that is essential to belief states that gives rise to this normative liability and warrants censure, we need some kind of explanation as to why that same intrinsic property of the belief states of non-agents does not give rise to the same range of normative properties for their beliefs. But that explanation seems difficult to provide. For instance, if the correctness condition is essential to belief so that belief is that mental state that is correct if and only if it is true, and the correctness condition is supposed to be a matter of what one ought to do *in that strong sense* in which adult humans ought not to have false beliefs, but the correctness condition can *also* be a matter of much weaker normativity, how can we be sure of the strong normativity for adult humans?

Without solutions for the sophistication problem and the censure problem, state doxastic normativism may carry the consequence that non-agents are excluded from counting as believers. For the sophistication problem, this would be because they do not seem capable of following certain epistemic norms, while the disposition to adhere to those norms is essential to belief. For the censure problem, the idea would be that since non-agents are not subject to strong belief norms, but strong belief norms follow from a normative element that is essential to the state of believing, their mental states are not beliefs. But many philosophers and scientists think that non-agents like animals and young children do have beliefs (see §3.1.2 for discussion). This is an example of a problem I call *overdiscriminate state exclusion*, which we will encounter several times throughout the thesis. Doxastic normativism's definition of belief risks being too narrow and excluding too many mental states from belief status.

I will not investigate in detail here the extent to which the normativist has to endorse state exclusion of non-agents. The sophistication problem seems like a serious worry for the dispositional argument, perhaps a kind of normativism not staked on dispositionalism could account for the discrepancy in epistemic normativity weighing on most humans and non-agents respectively. One solution, at least for the censure problem, would be to accept a much thinner notion of epistemic normativity, to the effect of denying that there is such a thing as real epistemic obligation that gives rise to the kind of censure we see in Ex. 3.3. Of course, that would be a serious theoretical cost, because we want epistemic normativity to yield exactly those kinds of judgments (for adult humans). Or the normativist could argue that the normative part of the nature of belief only founds some areas of epistemic normativity, including the judgments in Ex. 3.1, but does not directly generate the entirety of human epistemic obligation. Perhaps there are some additional premises that we can add to the (thin) normative nature of belief to yield that thick kind of epistemic normativity. I will not investigate these possibilities. Suffice it to conclude that without such additions to and clarifications of state doxastic normativism, the kind of normativity that is said to be essential to belief makes for difficulty for classifying many non-agential mental states as beliefs. For the rest of this section I will focus not on how the normativist could avoid non-agential state exclusion, but on why she should.

### §3.1.2 The argument from vestigiality

We will encounter more cases of overdiscriminate state exclusion throughout the thesis. Most of these cases will involve humans. Intuitions about whether some human subject has a belief or not in a given case are relatively clear. Such intuitions are much less clear about the kinds of non-agents we are concerned with in this section. I will focus on animal belief for what follows.

There is certainly controversy about animal belief, although I take it that machine belief is even more controversial. Philosophical arguments against the existence of animal belief have been around for decades (Davidson, 1982; Dennett, 1995; Stich, 1978), based on ideas like belief requiring concepts (Stich's argument) or belief requiring the concept of belief (Davidson's). In the contemporary literature, many defend the existence of beliefs in animals (Camp, 2009; Carruthers, 2004; Glock, 2010; Rescorla, 2009) but there is by no means a universal consensus. In light of this lack of universal consensus, one might think that the normativist can simply stipulate that non-human animals, or at least very many of them, do not count as believers on her view, and hence deny that the state exclusion in question is overdiscriminate. Nick Zangwill is a normativist who contemplates this possibility:

[I] can imagine someone arguing that animals and small children do not really have PAs. Perhaps we should require—as Kant would, I think, insist—that to have PAs we must conceive of ourselves as being under the sovereignty of reason. That is, we cannot have PAs without possessing epistemic concepts and having the capacity to reason. Perhaps one can only be bound by an obligation one is aware of. If so, there is no innocent virtue. On this view, there is no epistemic luck any more than there is moral luck. As far as Normative Essentialism is concerned, I can allow that one can have PAs without a more highfalutin conception of ourselves. So I can be friendly to animals and small children. But it could be that there is an independent argument for a more unfriendly Kantian conception of mind. In that case, Normative Essentialism could take that on board. (Zangwill, 2005, p. 16)

The idea seems to be that animal belief is an option the normativist can try to allow for, but that non-humans lacking beliefs is a cost she can accept as well. If we simply deny from the outset that animals have beliefs, then the state exclusion through the normative criterion we saw above will be to normativism's credit rather than an argument against it. But can we simply stipulate that animals don't have beliefs? I will try to argue in this and the following subsection

that that may be less straightforward than some normativists may think, and that it would not necessarily solve the censure problem anyway. My argument to this extent is not an a priori argument about the concept of belief. It takes certain empirical facts about humans explicitly as premises. Firstly, that we share vast amounts of evolutionary history and genetic material with many non-human animals. Secondly, that evolution tends not to eradicate biological structures very quickly.

Take as the first premise that there are non-human animals, ones that lack the sophistication required to count as rational on the above strict notion, that are conscious: they have mental states. To not beg the question against the sceptic about animal belief, we should not assume them to have *intentional* mental states, but their being conscious (which is fairly empirically uncontroversial (Low, 2012)) means they have some kinds of mental states.

As the second premise, take that some of these mental states are functionally similar to some of our beliefs. The dog has a perceptual input of hearing the dog food box being shaken and a behavioural output of running towards the sound of the shaking: highly plausibly some kind of mental state connects these two. We don't have to assume that mental state to be a belief or to have "food" as its content or anything like that, but in a conscious animal there is probably *something* mental that connects this perceptual input and behavioural output which we see connected reliably across many events of food box shaking. Let's call these mental states, (of which we are with the normativist assuming that they are not beliefs), "proto-beliefs".

Thirdly, if animals have these proto-beliefs that reliably connect their inputs and outputs, it is highly plausible that we have the capacity for those same mental states. This is highly plausible *even* if, with the normativist, we assume that humans have beliefs which fulfil that same

functional role. The reason is vestigiality. Biological structures tend to evolve away very slowly even after they have become redundant, the way our tailbone is redundant or even somewhat disadvantageous and the way our appendix is redundant and disadvantageous. We would expect proto-beliefs in humans to be redundant: we have beliefs which fulfil a very similar role, and presumably beliefs are better than proto-beliefs, hence us having evolved to have them. But since evolution dismantles biological structures slowly, we should assume that humans, too, have these mental states.

But also very plausible is the premiss that we *do not* have these proto-beliefs that aren't beliefs. We have no introspective evidence of them existing. While the normativist could argue to an extent that these states would be difficult to detect introspectively because of the simultaneous presence of beliefs, us having them and not being aware of them at all would be a mysterious result. The first result of the argument from vestigiality is then that animals are unlikely to have a different type of mental state instead of beliefs, for if they had them, we would be very likely to have them too, and we don't.

### §3.1.3 The norms of proto-beliefs

But the vestigiality argument makes for trouble even without the last premiss that contradicts human possession of proto-beliefs. Imagine that we really did have these mental states, the existence of which we are not aware of, in addition to our beliefs. They may be functionally redundant in the sense that we do not need them, but they would not be causally inert. Animals act on their proto-beliefs, and proto-beliefs might have the same effects on us. Even more clearly, the proto-beliefs would be likely to influence some of our other mental states, including our beliefs. Different types of mental states often influence one another (what we believe

influences what we fear, what we hope influences what we believe, what we estimate influences what we believe etc.). The proto-beliefs are often caused by perception, and many beliefs are caused by perception as well. So the proto-beliefs are functionally speaking not dissimilar to beliefs: their causal history is much like that of some beliefs and their causal powers extend to how we behave and what we think. In light of that, we would expect them to be subject to epistemic norms. Ascribing the kind of thin epistemic normativity to proto-beliefs that we saw in Ex. 3.1 should be unproblematic, for that normativity extends uncontroversially to animals as well. But say our cognitive neuroscience makes some quantum leaps and we're able to detect on a scan whether someone is in a given belief state or not, and we also understand the neurological difference between beliefs and proto-beliefs. Now we put Jack in the scanner because Jack poisoned his children with foxglove berries and we want to know why he did it. It turns out Jack lacked any positive beliefs about the edibility of foxglove berries (say the scanner can not only detect present mental states, but their history and origin as well), but we do find evidence of a proto-belief that foxglove berries are edible. What would we say about Jack? That he had a mistaken attitude but that since it was the kind of attitude animals also have he was not wrong to hold it, had no obligation to not hold this attitude? That seems counterintuitive.

Regardless of what the phenomenological differences between belief and proto-belief may be, I don't intuit that they would free Jack from epistemic normativity. Perhaps proto-beliefs are less vividly present in the mind than beliefs – but for dispositional beliefs the existence of which a subject is not even aware of (until she asks herself whether she believes the proposition in question) we don't intuit a lower amount of doxastic obligation. Perhaps proto-beliefs are much more difficult to control than beliefs – but again, for the states we currently assume ourselves to have there are many people who think doxastic obligation is compatible with involuntarism.

It is thus not clear that there is any reason why Jack's proto-belief should incur less censure than a belief he could have had instead, other than the fact that proto-beliefs are shared by animals. Furthermore, we don't have the intuition for other kinds of mental states that ones we share with animals are subject to less norms than ones unique to humans: animals can clearly experience fear, but potentially not jealousy, pride, or embarrassment. Yet if I am terrified of a harmless thing and jealous of a pitiable situation, the fault in those attitudes and the kinds of criticism I could incur seem the same.

It looks like epistemic norms would very possibly extend to human proto-beliefs if we had such states. But of course that replicates the problem that the move to deny animal belief was meant to evade: if the norms of belief are due to a normative element intrinsic to beliefs, the norms for proto-beliefs will presumably be due to an intrinsic normative element as well. But then how can it be the case that that intrinsic normative element makes for strong epistemic obligations for human proto-beliefs but not for animal proto-beliefs?

The problem the vestigiality argument turns on is that even though it might be plausible to stipulate that humans have the capacity for a certain mental state that animals are not capable of (after all there are many abilities and experiences unique to humans), it is much less plausible that all mammals and birds would have the capacity for something humans would lack<sup>29</sup>. Therefore, arguing that state normativism is compatible with weaker epistemic normativity for animals because animals are not capable of belief is not an easy fix for normativism. It opens

---

<sup>29</sup> Could we not just say that animals and humans share capacity for a single type of mental state, but that because of its increased functional potential in humans due to our increased cognitive capacity, they count as beliefs in humans but not in animals? That seems like a much more sensible strategy, but it is not open to normativists who argue that the normative criterion is part of the essence of belief states: the normativists are making claims about the nature of the mental state itself, not just about normative consequences that may follow from the role it plays in the subject's cognitive life at large.

up the task of explaining why humans should not also have proto-beliefs and what their normative status would be if we had them.

### §3.2 Dispositions and overdiscriminate definitions

Non-human animals' mental states are not the only ones that state normativism might exclude from counting as beliefs. Firstly, recall the premises from the argument from dispositions, which define beliefs as manifestations of certain dispositions which are rational. States that are the manifestation of irrational dispositions are explicitly excluded from belief status. As we saw in §2.4, this might seem initially plausible when we think of "irrational dispositions" as dispositions utterly devoid of rationality, but once we add the clarifications of the other premises of the argument from dispositions, it becomes clear that this requirement is actually a problematically strict one. Mental states resulting from the kinds of dispositions that the argument classes as irrational are categorised as non-beliefs, which is counterintuitive. As will become apparent over the course of the next three chapters, overdiscriminate state exclusion is a consequence of many versions of doxastic normativism widely construed.

The exclusion based on the wrong dispositions at the hands of the argument from dispositions is one of them. This problem for defining mental states through dispositions is well-known in the philosophy of mind where it has been a principal motivation for teleo-functional theories of the mind as a move away from traditional (dispositional) functional views. Disposition-functional views define mental states in terms of dispositions to produce certain outputs from certain inputs. The assumption behind that view is that what makes a mental state the kind of mental state it is, is how it functions. This is Wedgwood's view. Approaching this question from the philosophy of biology, teleo-functionalists reproach dispositional functionalism that it thereby

overlooks the possibility of *malfunctioning* mental states that behave very differently from other states of the same type. What makes a mental state the kind of mental state it is, on the teleo-functional view, is not how it does function but how it ought to function: what its function is. The importance of classifying mental categories by function as opposed to by functioning is especially salient against the backdrop of conceiving of them as biological categories. As Ruth Millikan (1984) writes:

The traditional functionalist has assumed that whatever interesting functions make beliefs to be beliefs, these must be among the actual dispositions or powers of beliefs or of people who have beliefs. ... [T]he traditional functionalist concentrates upon input and output and upon ways the system might work inside, hoping to show what about the insides of this system or about its dispositions constitutes its having beliefs. ... But if, instead, we push the analogy with biological categories, only *true* beliefs are capable of performing the defining function of beliefs. (p. 17-18)

The dispositions that all tokens of a biological category share are very few, according to Millikan: malfunctioning tokens are too different from functioning ones to yield a useful definition of the type they are both tokens of. Applied to belief, she is sceptical about the existence of an informative set of dispositions that both true and false beliefs share. That is why we should not expect much from dispositional-functionalist definitions of belief. The worry that using dispositions of *all* beliefs to define belief is uninformative carries with it the worry that an informative definition of belief that *does* use dispositions will not cover all beliefs. That is exactly what I will argue, and this time not only for potentially controversial candidate-believers like animals: if having the right dispositions is a necessary condition on being a belief, severely malfunctioning candidate-beliefs are not beliefs. I will work this objection out using the notion of norm guidance.

### §3.3 Norm guidance

To be guided by a norm means that some of one's behaviour is explained by the fact that one is subject to the norm. On the agential level, this usually takes on the form of conscious norm acceptance and endorsement and consequently using it as a premiss in one's practical reasoning. Crucially, this doesn't mean one always has to follow that norm. If I accept for myself the prudential norm "Don't drink coffee after 6", I will count as guided by it if it makes a difference to my decisions about when to drink coffee at least some of the time. It is compatible with being guided by the coffee norm to sometimes make exceptions. It is also compatible with being guided by the norm that one fails systematically in obtaining the prescribed result: if I avoid coffee after 6 most days deliberately because of my subscribing to the coffee norm, but my evil demon comes to pour coffee down my throat every night at 7, I will still count as being guided. So on the agential level, we have a relatively good intuitive understanding as to what guidance by a norm is. As we will see below, there is ongoing discussion in the literature about whether a truth norm is able to guide agents in this sense. But although my main opponent in this chapter is a notable defender of the truth norm, it is not the content of the norm of belief that is my main target, but its psycho-ontological status. Let's start with a brief overview of the importance that has been given to norm guidance in the literature.

#### §3.3.1 Content normativism and doxastic normativism

Content normativism is an application of meaning normativism, a theory from the philosophy of language that in its modern manifestation goes back to Kripke (1982). Meaning and content

normativism link meaning and content to the conditions of correct usage in some sense<sup>30</sup> or other. Several of the doxastic normativists we have encountered are normativists about content as well (Engel, 2000; Boghossian, 2003; Gibbard 1994). The two views also share critics, like Papineau (1999) and Hattiangadi (2006) along with Glüer and Wikforss themselves (Glüer & Wikforss, 2009, 2010, 2013b).

There are arguments for doxastic normativism that take the normativity of content or meaning as a starting point. This is one group of arguments for doxastic normativism against which I will not offer dedicated undercutting objections in this dissertation. I will here only note that both the truth of content normativism generally, as well as its bearing on the nature of belief have been subjected to much criticism in the literature already. Kathrin Glüer and Åsa Wikforss' work is among the most influential to that effect, through the powerful "no guidance argument" against the normativity of content, which translates into an argument against the truth norm of belief as well (Glüer & Wikforss, 2009). One premiss from that argument is that doxastic normativism involves norm guidance. Here is Glüer and Wikforss (Glüer & Wikforss, 2013a) on such guidance:

It is part and parcel of belief normativism as put forward in the debate that the rules or norms in question are genuine prescriptions capable of guiding belief formation. We shall argue that this is very hard to square with some basic intuitions about rule guidance. Any account of rule-guidance needs to support the distinction between being guided by a rule and merely being in accord with it. The rule must make a difference, so to speak, a difference both to the formation and, consequently, to the explanation

---

<sup>30</sup> The main distinction among these senses has to do with the question which determines which. Glüer and Wikforss (2009) draw a distinction between meaning engendered normativity and meaning determining normativity, where the former would mean that meaning gives rise to norms and the latter would mean that norms give rise to meaning.

of rule-guided belief. A rule-guided belief is a belief that, in a sense to be further specified, has been formed because of the rule. (p. 123)

The no-guidance argument is designed against deontic versions of doxastic normativism that make a truth norm essential to belief. Such accounts, Glüer and Wikforss write, need to provide a story as to how the truth norm guides belief formation. But in order to follow the rule “believe p only if p is true”, the subject needs to form a belief about the truth of p, making the rule impossible to be guided by due to infinite regress.

Glüer and Wikforss focus on prescriptive norms, but on doxastic normativism widely construed as I have defined it, it is not the case that “[i]t is part and parcel of belief normativism as put forward in the debate that the rules or norms in question are genuine prescriptions”. Definitions of belief that make evaluative norms part of the definition of belief count as normativist just as much. And there are self-styled normativists like Fassio (2011, forthcoming) and Nolfi (2015) who take non-prescriptive norms to be constitutive of belief. But despite this taxonomical disagreement with Glüer and Wikforss, I agree with them in spirit: the guidance requirement makes for trouble for doxastic normativism across the board.

### §3.3.2 Intrinsic norms and guidance

State doxastic normativism is a theory about the nature of one of our mental states. As such, it comes with a commitment to a kind of norm guidance that goes further than the agential notion above of using an endorsed norm as a premiss in one’s practical reasoning: on state normativism, the condition for correctness is written into the mental states themselves in the form of dispositions. It is by virtue of these dispositions that norm-governed mental states display a certain sensitivity to factors that are relevant to their essential norm. Beliefs, for

instance, seem sensitive to evidence quite strongly, to the extent that it is very difficult to form them with non-evidential data as input<sup>31</sup>. That belief has these descriptive properties of being sensitive to certain factors that are relevant to epistemic norms has been used as evidence for its intrinsic normativity for decades. So there seems to be a strong connection assumed between the intrinsic normativity of a mental state and the tendencies of that state to behave in norm-appropriate ways. Of course, on state normativism that involves reference to dispositions in its typing of belief, that is true by definition. But we see authors across the normativist spectrum, including ones that make no mention of dispositions to believe in their accounts, take the power of their preferred belief norm to guide the formation of actual beliefs as a key premiss to defend, and the causal relevance of the intrinsic belief norm is a feature in some sense or other in all the accounts we have considered or will consider.

Shah and Velleman's conceptual normativism includes an account of being moved through norm acceptance (Shah, 2003; Shah & Velleman, 2005), which one of them has explicitly defended against criticism (Shah, 2013). Steglich-Petersen, too, has devoted work explicitly to defending the claim that his account of the belief norm can account for guidance (Steglich-Petersen, 2010). Then there are of course Wedgwood's (2007b, 2009), Nolfi's (2015), and Zangwill's (1998b, 2005a) accounts of state doxastic normativism that define belief in terms of dispositions to form and revise beliefs in certain rational ways, meaning candidate-beliefs only count as beliefs if their causal history is at least minimally norm-guided. Further examples of

---

<sup>31</sup> The relationship between belief and evidence is discussed at length in §4.2 and its relevance for doxastic freedom in §7.2.

normativists that explicitly defend guidance through intrinsic norms include Engel (2013a, 2013b), Toppinen (2015) and Boghossian (2008).

To describe this common tendency among doxastic normativists, which I will argue makes for a key difficulty for all these accounts, I will use the term *norm effectiveness*. Norm effectiveness captures the causal power of a norm to make a difference in the formation and use of those things it ranges over. Norm effectiveness is not at all a necessary requirement on the truth of a norm: “Keep your seatbelt fastened until the aircraft has come to a complete standstill” is a highly ineffective norm in the sense that many people make no attempts to follow it, but that datum does not make us doubt that the seatbelt norm exists. For the intrinsic norms that the normativist is interested in, matters are a little different. Minimal norm effectiveness is assumed, and defended at length<sup>32</sup>, and there are general reasons to expect norms that are essential to a certain behaviour to guide that behaviour at least to an extent.

First of all, the norm is by definition in place whenever the behaviour takes place; there is no opting out possible. Furthermore, it seems to be a common intuition that someone who behaves in a way entirely unguided by the rules intrinsic to a given practice is simply not performing that practice. Games are a common example. If every candidate-move of mine on a chess board violates the rules of chess, to many people it will seem intuitively clear that I am not playing chess poorly but that I am not playing chess at all. Another reason to expect accounts of intrinsic norms to feature some sort of norm guidance is evidential: even though

---

<sup>32</sup> See for instance the exchange between Wedgwood and Georges Rey in (B. P. McLaughlin & Cohen, 2007) about the amount of irrationality that is compatible with Wedgwood’s normativism.

subjects demonstrating guidance by a given norm is only partial<sup>33</sup> evidence for the existence of that norm, it is, apart from intuition, some of the only evidence we have for it. Depending on one's meta-normative view, there is nothing incoherent about the idea of an intrinsic norm that has no effectiveness (no one follows it either in intention or in result). On the first-order normative level however, we would expect discussion of such entirely ineffective norms to be rare, because of the evidence question as well as because such causally inert norms might just be of less interest to human inquiry: arguing at length for the inclusion of some element in the description of the nature of a given thing if this element will be causally inert on the descriptive level seems pointless. I cannot go into much more detail here about what norm effectiveness is and how it works, because it is an artificial term meant to denote a relatively disparate disjunction of theoretical elements of different versions of doxastic normativism. On conceptual normativism, effectiveness is supposed to be delivered through the way subjects are moved in their behaviour by the norms they consciously endorse. On teleo-functionalism, effectiveness comes from functioning mental states, where the expectation is that evolution makes for a tendency towards functioning mental states. On teleologism and state normativism, a certain minimal tendency to conform to epistemic norms is built into the nature of individual belief states, in the latter case through appeal to the dispositions that cause them.

---

<sup>33</sup> The debate between defenders of the truth norm of belief and the knowledge norm of belief is indicative of this inconclusiveness. It seems clear that most epistemic agents follow one of these two norms most of the time, but from that behaviour it is not obvious which of the two is the one making for the relevant guidance: many instances of attempting to have one's beliefs be true might actually be instances of attempting one's beliefs be known, and many instances of adhering to the knowledge norm might actually be instances of someone trying to adhere merely to the truth norm but being very careful about it. Furthermore, even when it is clear what the content of a given norm is in light of subjects' behaviour in response to that norm, it might still be unclear what kind of norm it is: the distinction between religious, cultural, moral, and aesthetic norms is often difficult to tell make based on their guidance over people.

The objection I develop in the rest of this chapter, which I call the *lawless deviance objection*, focuses on the relationship between intrinsic epistemic norms and norm effectiveness. We will ultimately encounter versions of this objection for many different formulations of normativism, and every time the assumption of some sort of causal effect of the epistemic norm on the way beliefs tend to be in reality will prove to be the source of the problem. For the current presentation of the objection, the normative dispositions element of state normativism is the relevant manifestation of norm effectiveness. On the normative dispositionalism picture, belief formation is by definition guided by the epistemic norm, for that norm finds its origin in the dispositions to form beliefs rationally, and something is only a belief if it is a product of these rational dispositions.

### §3.4 The lawless deviance objection

On normative dispositionalism like Wedgwood's, it is not impossible for *some* irrational dispositions to be at work in the formation or maintenance of an individual belief. It does need to be the case that there are also some rational dispositions however. How to understand that in detail is a question I will seek to answer below. But for now let's note that there has to be some extent to which a belief is caused by rational dispositions, and hence norm-guidedly formed, or else it will not count as a belief on state normativism: state normativism requires a certain amount of norm effectiveness. The first step for my lawless deviance argument is the observation that being caused by rational dispositions, being norm guidedly formed<sup>34</sup>, comes

---

<sup>34</sup> Of course, not only the formation of a belief needs to be guided by norms or the product of rational dispositions, but its maintenance does too. I use "norm-guidedly formed" as short-hand for "formed and maintained under guidance by epistemic norms".

in degrees. If being guided by epistemic norms is a matter of manifesting some rational dispositions, more or less of these dispositions can be manifested.

### §3.4.1 Believers and beliefs

At this point the normativist might interject that although some believers are more guided than others, all believers are guided *tout court*. She could claim that there is no such thing as being guided by epistemic norms to such a minimal extent that one does not count as a norm-guided agent at all. Is this true? If being guided by epistemic norms *tout court* is simply a matter of *sometimes* adopting beliefs the formation of which was guided by some norms, then I would be inclined to agree with the normativist that all creatures that we should count as believers will qualify as such by the being-guided condition. There might be hypothetical creatures that strike us as believers but that fail the condition of ever being guided by epistemic norms due to some kind of evil demon scenario. But we won't need to engage with that question for present purposes. Because, crucially, doxastic normativism is not a theory about the ability to believe, but a theory about belief.<sup>35</sup> Indeed, normativists explicitly appeal to the norm of belief as a demarcating criterion between beliefs and non-beliefs. It is a condition for belief that its formation is the product of rational dispositions, and mental states whose formation is not a product of the right dispositions are not beliefs. As we have seen in §2.4, this makes for the possibility of overdiscriminate state exclusion: some plausible candidate-beliefs might be excluded by the rational formation criterion. The objection I am currently developing builds on overdiscriminate state exclusion, yet it goes further than that argument because it depends less

---

<sup>35</sup> Although the conditions for being a believer are sometimes appealed to in order to argue for normative features of belief (Nolfi, 2015).

on intuitions about which states count as beliefs (and hence which exclusions are overdiscriminate). It instead delivers a reason to think that the excluded states in question must be beliefs.

### §3.4.2 Degrees of guidedness

The premiss I am working with is that, on the level of individual mental states, some are more guided by the epistemic norm than others, because they are the product of more rational dispositions. Let's use an example. Take a restrictive truth norm "Believe  $p$  only if  $p$ ". The formation and maintenance of a belief will count as norm-guided if it is the product of dispositions that take only truth-relevant data as their input condition <sup>36</sup> (for the next two subsections, "norm-guided" refers to being guided by *this* norm). The idea is that we have some dispositions to believe  $p$  based on input that counts as truth-relevant (perception, testimony, calculation), but potentially also some dispositions to believe  $p$  based on input that is not truth-relevant (e.g.  $p$ 's content being good news). Now a belief presumably counts as norm-guidedly formed if it is the consequence of the manifestations of dispositions that take such truth-relevant data as their input.

Ex. 3.4: I am reasonably good at counting, which means that I have a disposition to form the belief "there are  $n$   $X$ " after counting the  $X$  that, more often than not<sup>37</sup>, manifests when there are  $n$   $X$ . Now I form a belief that there are 20 marbles as a consequence of counting 20 marbles, and no other dispositions to believe that there are  $n$   $X$  (such as a disposition to always go for 20, or to find even numbers more likely, or to match  $n$  to

---

<sup>36</sup> On pain of triviality, "data" is used non-factively here.

<sup>37</sup> Whatever the threshold of reliability is.

the colour of the objects one is counting,...) play a causal role in the formation of my belief that there are 20 marbles.

The belief that there are 20 marbles in Ex. 3.4. is norm-guidedly formed, since it is entirely the consequence of the manifestation of a rational disposition. Many beliefs are more complex in their formation.

Ex. 3.5: I am not very good at counting. I have an average error rate of 3% when counting things in excess of 100. I know this about myself. So I have a disposition to believe that there are  $n$  X after counting the X that manifests when there are  $n$  X less often than above because its manifestation conditions are less truth-relevant. However, I also have a disposition to lower my confidence in the results of my counting that truth-relevantly takes my belief about my error margin as its input condition and I have a disposition to translate my confidence levels in  $n$  into approximations of  $n$ . As a consequence of these dispositions manifesting, and no others manifesting, I count 167 marbles and form the belief that there are at least 160 marbles.

I would assume that the belief that there are at least 167 marbles in Ex. 3.5. also counts as norm-guidedly formed. With these illustrations in place, we can now see that, since belief formation is often the result of multiple dispositions, and since dispositions can be more or less truth-relevant, being norm-guidedly formed comes in degrees. The more good dispositions and the less bad dispositions are causally responsible for a given belief, the more we can speak of that belief as norm-guidedly formed. The question now is how to translate this graded notion

of norm-guidedness into the categorical notion of being norm-guided that is supposed to be a make for norm effectiveness on state normativism.<sup>38</sup>

The first option would be to discount any belief that has dispositions in its causal chains that take non-truth-relevant data as their input condition. If this were the notion of norm-guided formation at hand, my objection would be a very easy one to make and a virtual death-knell to doxastic normativism: vast amounts of our beliefs have *something* in their causal history that is not truth-guided in the sense we're talking about: biases increase our confidence in otherwise well-formed convictions, questionable heuristics speed up our reasoning, little mistakes creep into complex mental operations. Discounting all of these from qualifying as beliefs (due to classing them as not norm-guidedly formed) would be deeply implausible. However, I think it would be fairer to the normativists to interpret her notion of norm-guidedness as less forbidding. Take the following kind of case.

Ex. 3.6: Molly forms the belief that her husband has been coughing in his sleep for 20 nights in a row. She counted 20 different consecutive nights in which she remembers hearing him coughing. However, one of those 20 nights, Molly was actually asleep and dreamt that she heard him coughing instead of hearing him coughing.

Let's stipulate that for the subject in Ex. 3.6, the results of counting are truth-relevant input for belief, as are products of memory for the kind of timespan at hand, and that dreaming does not produce truth-relevant input for mental states. Surely, we would not want to say that the formation of this belief (which is made up of 19 instances of a veridical perception disposition,

---

<sup>38</sup> I have not addressed the question whether intrinsic normativity itself might also come in degrees. But if that were the case, the threshold question of norm-guidedness would persist, albeit with the qualification that its answer should be compatible with whatever the intrinsic normativity threshold is.

remembering, counting until 20, and one dream) is entirely un-guided by epistemic norms. I can count as steering my car even though 1% of its actual movements are determined by bad mechanical processes in the car. Now take another case:

Ex. 3.7: Molly forms the belief that her husband has been coughing in his sleep for 20 consecutive nights. The following input conditions brought this belief about. She hears her husband coughing in his sleep right now and is very irked by this. She remembers reading that when nightly coughing persists for more than 19 nights one should go to the doctor about it. She also falsely believes that were her husband to go to the doctor tomorrow, the irksome coughing would stop immediately. Molly doesn't remember much about whether coughing happened the previous 19 nights.

Let's stipulate that for the subject in Ex. 3.7, the result of hearing is a truth-relevant input for belief, and being irked by something is not truth-relevant input, nor is the disposition to take the false belief that going to the doctor will prevent the husband from coughing again as input for how often he has been coughing. The belief in Ex. 3.7 that the husband has been coughing for 20 nights is caused partially by dispositions with truth-relevant input (on one of the 20 nights, i.e. right now, Molly heard him coughing) and in other, large parts by dispositions with bad input. Clearly this belief is less norm-guidedly formed than the one in Ex. 3.6: the proportion of rational to irrational dispositions that caused it is lower. It seems to me that this belief will indeed not count as norm-guidedly formed in the categorical sense: too few of the dispositions that caused it are rational, the norm-guided part of its formation was too minimal.

#### §3.4.4 A minimal threshold?

One might think that Ex. 3.7. nevertheless counts as norm-guidedly formed because one thinks that *any* cognition that has *any* good dispositions whatsoever in its causal history counts as norm-guidedly formed. But consider that even more spectacularly bad beliefs than Ex. 3.7 often have *some* truth-relevant input. Take a case of hallucination:

Ex. 3.8: Rex believes there is a dinosaur in their room when there is nothing resembling a dinosaur to be seen. Lots of bad dispositions are at play here: one that takes imagination instead of perception as an input for a perceptual belief, and multiple dispositions that would have used evidence that the dinosaur cannot be real to counteract the perceptual pull to believe (like dinosaurs can't open doors without breaking them, Rex has good reason to think dinosaurs are extinct etc) fail to manifest. But imagine that yesterday Rex also hallucinated the dinosaur, and hallucinated that they had a fight with it and injured it, and today it is part of their belief about the dinosaur that it has some dried blood on its skin.

Again, we have a mental state caused and sustained predominantly by bad dispositions manifesting and good dispositions failing to manifest. But one element of the hallucination comes from a rational disposition: the disposition that takes the memory of seeing an injury yesterday as input and the belief that the injury will still be there today as output. Even deeply disturbed cognitions can be partially determined by rational dispositions. I think it is clear though, that we do not want to call the formation of Rex' belief "there is a dinosaur in my room who got here by opening the door and who has dried blood on its skin" guided by the truth norm, just like we would not count an aquaplaning car as steered by me even if my

manipulations of the steering wheel still have a small effect on the swerving pattern. If the only way for a cognition to not count as norm-guidedly formed is for it to have no rational dispositions in its causal history whatsoever, the notion of norm-guidedness would be highly trivialised and the emphasis given to accounting for it with our epistemic norms would be mysterious. So I conclude that somewhere between Ex. 3.4 and Ex. 3.8 we want to draw the line of what counts as norm-guidedly formed, where I would put Ex. 3.5 on the side of norm-guidedly formed and Ex. 3.6 on the other, but where intuitions about the exact cut-off point might well differ.

#### §3.4.5 Lawless deviance

Recall that on state normativism, being caused by rational dispositions and hence being norm-guidedly formed is a necessary condition on belief. We can see, then, that the thesis that beliefs are essentially normative, where this is spelled out in a way that assumes norm effectiveness, makes for the disqualification of certain mental states as beliefs on the grounds that their formation failed to be guided by epistemic norms. The candidate-beliefs in Ex. 3.7 and Ex. 3.8 are not beliefs on this view.

Doxastic normativists are not unaware of the possibility that their accounts entail some unintuitive state exclusions; Nishi Shah (2003), for instance, discusses as a problematic element of teleologism that it cannot count wishful thinking (where the subject fails to aim at the truth in the right way) as belief. Now perhaps normativists could accept the theoretical cost of these state exclusions, in the same vein in which Nick Zangwill considered the possibility of excluding animal belief. The cases above were intended to be examples where the intuitive pull of classifying the mental state in question as a belief is strong, and where consequently its

disqualification appears implausible. Nevertheless, the move of taking these disqualifications on board, albeit a very costly one, remains open to the normativist. This is where lawless deviance comes in.

Doxastic normativism is a theory about the nature of belief and a theory about the source of the epistemic norms that govern belief. Particularly the latter dimension is often hailed as a motivation for normativism. Recall Kate Nolfi's (2015) quote that: "[n]ormativism enjoys a great deal of popularity in epistemology because it promises to explain, in a relatively straightforward way, the authority of epistemic norms: on any version of normativism, beliefs will be appropriately evaluated with respect to epistemic norms simply by virtue of being the sorts of mental attitudes that they are." (p.181). Christine Korsgaard (2009) writes, about constitutivism generally: "[t]he idea of a constitutive standard is an important one, for constitutive standards meet skeptical challenges to their authority with ease." (p.29) Normativism is supposed to help us defend our fundamental commitment to the existence of epistemic norms against the normative sceptic. The sceptic looks unmoved at our idea of the existence of an epistemic norm that weighs on people's beliefs, that makes it the case that there are things one should and should not believe, that some beliefs are better than others. Why, asks the normative sceptic, should she care about the content of these so-called norms? One of the supposed selling points of normativism is that it safeguards the validity of epistemic censure.

Lawless deviance threatens exactly that. We need just two premises: 1) Some candidate-beliefs are not beliefs because their formation is not appropriately norm-guided. 2) Because they are not beliefs they are not subject to the norms of belief. In the conclusion, we have ended up with a group of cognitions that seem deserving of serious epistemic censure but fall completely

outside of the scope of the norms we have been working to defend to do this kind of censuring with. What's worse, is that those states that are excluded from censure by the norms of belief seem to be among those we would want to condemn the most: convictions formed as a consequence of delusions, wishful thinking, bias, bad heuristics, haste.

Its application to the most problematic cases and its ability to denounce them is an acid test for any norm. If there is such a thing as a sufficient level of deviance above which the very norm that censures this kind of deviance (below the threshold) ceases to apply, that seems like a serious indication that the norm does not function correctly, does not take the right criteria as its input and hence that its verdicts across the board are suspect. The following application of the principle of lawless deviance should drive home that problem.

Ex. 3.9: Tom is a man from an ethnic group that Mariah and Betty are both biased against for reasons that stem from racism and are entirely truth-irrelevant. It is part of their bias that they are disposed to think members of the group in question are liars. Tom is not a liar. However, he unfortunately, for idiosyncratic physiological reasons that have no bearing on his sincerity, often looks and behaves like a typical liar would (he sweats a lot, has a tendency to look around him shiftily when telling a story, often tells stories so that they sound incoherent, tends to react startled when asked probing questions etc.). Now Tom makes a true assertion to Mariah and Betty. Mariah used to be a detective and is very acquainted with the physical signs of lying. Betty is not. Betty is however much more racist than Mariah is. Mariah and Betty both form the conviction that Tom is lying. Mariah does so mostly because of Tom's behaviour, which she mistakenly interprets as showing him to be lying, but her confidence is bolstered by the

bias that disposes her to think Tom more likely to be a liar than the average person.

Betty thinks Tom is lying mostly because of her racist bias towards him, although his strange behaviour increases her confidence somewhat.

Say that Mariah's belief is the product of 80% truth-relevant dispositions and 20% truth-irrelevant ones, and that Betty's is the other way around, and that the threshold for categorical norm-guidedness is somewhere between Mariah and Betty. Now Mariah's attitude counts as a belief and hence gets evaluated by belief norms: it gets censured for being false, and for having been partially grounded in poor evidence. Betty's attitude however falls below the norm-guidedness threshold necessary to count an attitude as intrinsically normative and hence as a belief, and since it is not a belief the norm of belief does not apply to her. The same deviant disposition that Mariah's belief gets censured for goes uncondemned in Betty's case, despite the deviant disposition being much stronger for her.

### §3.4.7 State-differentiation

As we will see throughout the next three chapters, the lawless deviance objection can be brought to bear against all versions of doxastic normativism widely construed that involve a norm effectiveness requirement. Now one general feature of the over-exclusion scenarios I discussed, is that we don't seem to have a very intuitive grip on whatever falls on the excluded side of doxastic classification. Say we decide that the subjects in Ex. 3.7, Ex. 3.8, and Ex. 3.9 do not have a belief. How ought we to classify their attitude instead? Our existing mental state concepts don't seem to account for these phenomena, they don't seem to fit any of the other affirmative attitudes. Compare this with the way we think about other practices that are thought

to be constitutively normative, take playing chess and making assertions<sup>39</sup> as examples. Something is only a chess move if it is governed by the rules of chess. If I randomly move some pieces around on a chess board, I am not making chess moves *even if* those chess piece manipulations are physically exactly like real chess moves would have been. Something is only an assertion if it is governed by the norm of assertion. If I utter the words “The doctor and Philip had come over the road” out loud in the context of learning Ashbery’s “The Tennis Court Oath” by heart, I am not making an assertion even if the sounds I make are physically exactly like an assertion would have been. All of this is clear, and it is so partially because it is easy for us to understand what the person with the chess board or the person with the Ashbery anthology are doing: they are moving some chess pieces around and saying some words. But what is the person doing who has a mental state that is like a belief descriptively but isn’t a belief for normative reasons? That is much less clear.

Of course, a lack of an existing concept in our theoretical landscape of thinking about the mind does not immediately indicate that nothing meaningful could fill that gap. As the literatures on aliefs, credences, and dispositional beliefs make obvious, new concepts for mental states we did not previously have the tools to think about can make for very illuminating additions to our theories. Let’s imagine that this concept-building challenge can be met, and that the excluded candidate-beliefs do form kinds that we can fruitfully label. For that kind of mental state, which as we saw faces lawlessness from the belief norms, the question now arises what its normative status might be *if not that occupied by belief* (i.e. being subject to the norms of belief). For if

---

<sup>39</sup> For the reader who wonders about my general stance on norm constitution: I think very good cases exist for the constitutive normativity of games and speech acts like assertion. But assertion is different from belief for one, importantly, because it is inherently social, and hence it is much more plausible that it has a culturally constructed element. I think of belief in contrast as a natural kind.

the doxastic normativist can show that although the deviant states in question cannot be censured by the norms of belief, there is some other affirmative state norm that can perform the censuring work in question instead, the lawless deviance objection will lose much of its sting. In light of that, offering an account of these alternative affirmative state norms is exactly what the normativist should do to make her theory of belief more plausible. I will finish my discussion of the lawless deviance objection with some observations on the difficulties facing the normativist in this task.

Tackling the question which norms govern the affirmative states that are not beliefs is challenging firstly simply because of the amount of work it opens up. Philosophers are already having much difficulty reliably identifying and satisfyingly formulating a norm for belief, and it is clear that we are very far away from having norms for guessing, estimating, suspecting, assuming, and speculating. The same goes for all the kinds of mental states that are classed as beliefs now but that are disqualified by doxastic normativism. For the doxastic normativist, the problem is particularly bad. First of all, on many views, being intrinsically norm-governed is supposed to be a special thing about belief that helps us differentiate it from other mental states. The need to stipulate norms for these alternative affirmative states undermines the state-differentiating motivation for normativism: instead of allowing us to demarcate belief as *the* mental state that is intrinsically epistemically normative, we would have to demarcate it by the *kind* of intrinsic epistemic norm that governs it.

Ralph Wedgwood's position falls into that category: state-differentiation, on his view, is achieved by the content of the specific norms of mental states, and intrinsic normativity is a property of the intentional generally. So belief differs from its affirmative kin because of the kind of norm

governing it. We have already encountered a problem for this strategy in the form of the dilemma from correctness conditions in §2.3. We will return at length to the task of state-differentiation in Chapter 4. But now we can see another difficulty for the thesis that the intentional itself is normative. Intentional states are meant to have state-essential norms on this thesis. These norms are supposed to display some amount of guidance over the mental states in question. If a token mental state's causal history is not norm-guided, the essential normativity of that type of mental state becomes much less plausible. But the kind of mental state the lawless deviance objection turns on, for which we need a norm to be in line with the thesis that the intentional is normative, are characterised exactly by their *lack* of norm-guided formation. Therefore, identifying any kind of intrinsic norm for these maximally deviant states will be difficult, unless we give up the idea that essential state norms display norm effectiveness over mental state formation. But that would be difficult to argue: why should the norm intrinsic to belief have the power to guide the formation of beliefs, but the norm intrinsic to the deviant states not have the power to guide their formation? Any kind of normativism that puts emphasis on the power of the intrinsic belief norm to guide belief will have to account for the deviance of the worst mental states falling into lawless territory. The deviant states that fall outside of what normativism defines as belief, are thus not only lawless by the lights of the norm of belief, but by the lights of epistemic normativity generally.

### §3.5 Against state normativism

This chapter discussed two lines of argument against state doxastic normativism. The first had to do with non-agential belief. Many non-agents seem less cognitively sophisticated than the adult humans that typically inspire accounts of belief. If the norm of belief is meant to explain

all our normative judgments about adult humans, such as the impermissibility of defeater-insensitive belief-formation, but is also constitutive of belief, non-sophisticated candidate-believers may be excluded from belief. Secondly, even for judgments through epistemic norms minimal adherence to which does not necessarily require much cognitive sophistication, such as the truth norm, non-agential beliefs may be problematic. Norms like the truth norm seem to apply to non-agents with much thinner normativity, as borne out by the kind of censure that seems appropriate for them. State normativism will thus either have to downgrade the normativity that is supposedly intrinsic to belief to something that is applicable to non-agents, or it will have to accept widespread state exclusion for non-agential belief. We looked into the latter strategy as applied to animals, and I argued that it is difficult to coherently establish: denying animals belief would mean they very probably have a different kind of mental state instead, but vestigiality makes it unlikely that animals would have a cognitive structure that we lack. As a consequence, questions about what our mental life has in common with that of animals, and what that means for the norms ranging over theirs and ours respectively are not avoided simply by denying that animals have beliefs.

The second line of argument I developed drew on the norm effectiveness that, for good reason, is part of most normativist accounts. If the essential norm is expected to have effects for belief formation and maintenance, problematic consequences follow for mental states that do not seem regulated or guided by any kind of epistemic norm. I called this problematic consequence “lawless deviance”: if we allow for state exclusion by normative criteria (token state  $x$  is so epistemically deviant it does not count as a belief), those excluded states will also be excluded from censure by the norms they failed to conform to. In other words, we can’t speak of them “failing” to conform to that norm in any robust sense at all. Glüer and Wikforss (2013b) write:

"It is a good question how irrational a state can be and yet be classified as a belief. Even if a plausible account of belief must allow for the possibility of irrationality, intuitively there must be some limits to the irrationality allowed for the state to qualify as belief." (p. 147). Much of this thesis is an answer to this question: even if one shares the intuition that there are limits to how irrational beliefs can be, we may nevertheless not want to adopt such limits in our definition of belief. For if something is not a belief, it is not governed by the norm of belief. If it is too irrational to be governed by the norm of belief, by what lights are we calling it irrational?

## Chapter 4: Against Conceptual Normativism

### The transparency of doxastic deliberation, and the challenges of state-differentiation

Having concluded against normativism (narrowly construed) about the state of belief, in this section I consider the view that the concept of belief is a normative one (hereafter “conceptual normativism”, short for “conceptual doxastic normativism narrowly construed”). I will focus on the version of this view on which, in order to have the concept of belief, and thus be capable of reasoning about belief, one must grasp the norm of belief. This is the version of doxastic normativism that is put forward by Nishi Shah and defended through his argument from transparency (Shah, 2003; Shah & Velleman, 2005). Pascal Engel, another prominent defender of conceptual normativism, appeals to the argument from transparency as well (Engel, 2013b). There are ways to make conceptual normativism a stronger thesis by linking it to claims about the nature of belief. Davidson, for instance, argued that possessing the concept of belief is a necessary condition for possessing the attitude of belief (Davidson, 1975). Shah and Velleman’s view, as we will see, is more nuanced, and would i.a. seem to allow for creatures that do not possess concepts to be capable of belief. It is only the weaker claim, that the concept of belief is normative, that I will explicitly address, and not the stronger claim that possession of this concept is necessary for possession of a mental state. It is however clearly necessary to possess the concept of belief in order to *think about* belief, and that is one consequence of Shah and Velleman’s argument that I will be concerned with. I will proceed as follows. First (§4.1), I give an overview of Shah and Velleman’s argument for conceptual normativism. Secondly (§4.2), I go over some issues with their argument to undercut the case for conceptual normativism,

before turning to (§4.3) some rebutting considerations to their view. I finish the chapter on some observations about state-differentiation (§4.4).

#### §4.1 The argument from transparency

Shah and Velleman's argument from transparency is, along with the argument from dispositions and the argument from correctness conditions, one of the only positive arguments for doxastic normativism put forward in the literature, and it is the main one for conceptual doxastic normativism. On their view of belief, belief can result from sub-intentional processes (like perception) or from what they call "doxastic deliberation", a conscious decision to believe that involves asking oneself the question "whether to believe that p?". This latter kind of belief formation, according to the authors, displays a specific psychological phenomenon they call "transparency":

The feature that we call transparency is this: The deliberative question whether to believe that p inevitably gives way to the factual question whether p, because the answer to the latter question will determine the answer to the former. That is, the only way to answer the question whether to believe that p is to answer the question whether p. (Shah & Velleman, 2005, p. 499)

The argument from transparency is an inference to the best explanation. Transparency, Shah and Velleman argue, is a robust feature of belief formation that our epistemological theory needs to provide some kind of explanation for. Furthermore, transparency marks a significant difference between belief and other affirmative states:

[Transparency] serves to differentiate [belief] from the other cognitive attitudes... [T]he only way to answer the question whether to believe that p is to answer the question whether p. By contrast, the answer to the question whether p will not settle either the question whether to suppose that p nor the question whether to imagine that p, and so

those questions do not give way to it-or, as we shall say, are not transparent to it. (p. 499)

I call these the two explananda of the argument from transparency: firstly, that belief deliberation is transparent, secondly, that belief is the only affirmative attitude the deliberation of which is transparent.

#### §4.1.1 Teleologism and transparency

Why should we need the normativity of the concept of belief to explain the transparency of doxastic deliberation? To answer this, Shah and Velleman consider at some length the way their rival view, teleologism, would go about the explanandum of transparent doxastic deliberation. Teleologism, which the authors take to be one of the main alternatives to doxastic normativism, is the view that belief has an intrinsic aim, which can either be construed on the level of the conscious intentions held by the subject (Steglich-Petersen, 2006), or on the level of sub-intentional cognitive mechanisms of belief production (Velleman, 2000b). The notion of belief's aim goes back to Bernard Williams (1973). Williams wrote that it is "a characteristic of beliefs that they aim at the truth" (p. 148), which he saw as the explanation for the constraints on what we can believe: we cannot bring it about on nothing but our own will that we believe something. Many authors in the belief norm literature accept some version of the claim that belief aims at the truth. The crucial difference between teleologists and normativists is that the former see the truth aim as an intrinsic property of belief (belief has a built-in goal, much like aiming at a target does), whereas the latter see it as a metaphor that is to be cashed out in normative terms (beliefs ought to be true, therefore the believer has a normative reason to strive for truth). Chapter 5 is devoted to a detailed investigation of the merits and shortfalls of teleologism. For now, let us just focus on its stance on transparency.

The transparency principle is meant to capture a phenomenon of evidential sensitivity. We cannot use just any kind of reason when trying to make up our mind about what to believe, there is a significant force exercised by evidential reasons. And evidential reasons are exactly those things that are relevant to answering the question "whether p"? One might think that the teleological view is ideally equipped to explain psychological phenomena of evidential sensitivity like transparency. Consider an obvious example of a goal-directed activity like taking a shot (at a target). There are reasons I can employ to shoot a certain way that are relevant to the reaching of the target: the distance of the target, my shooting skills, the wind, the type of gun I'm using etc. There are also reasons that can factor into my way of shooting that are not relevant to the reaching of the target, like how tiring they are for me or how cool they look. The goal-directedness of the act of taking a shot puts constraints on the amount of weight that the target-irrelevant factors can have in determining the way in which I shoot. If I decide to shoot with the gun pointed towards the floor because my hand is presently hanging down and shooting in that position will be less effortful than lifting my arm to point at the target that is located on the other side of the room, I simply won't count as taking a shot at the target. Belief aims at the truth in virtue of its nature according to teleologism, which can be expected to put constraints on the weight that can be given to truth-irrelevant reasons (like how comforting, or costly, or useful a given belief would be) in deciding what to believe. The phenomenon of transparency, whereby the question "should I believe p?" is answered conclusively by one's answer to the question "whether p?", is such a constraint. Hence, teleologism would seem to have resources for a plausible explanation of this phenomenon.

According to Shah and Velleman however, the teleological account of transparency ultimately fails. To establish this, the first step in their argument is to explain why teleologism cannot be

understood as *entirely* about conscious intentions. Even if purposive aiming at truth is necessary for some kinds of belief, they claim that teleologists will also have to situate truth-aiming mechanisms on the sub-intentional level. Shah and Velleman insist that, while there are some cases of belief that at least seem to be the consequence of a goal-directed deliberative process, many, if not the majority of, other beliefs are caused by processes that cannot be said to involve aims or goals, such as perception. Therefore:

If the metaphor that belief aims at the truth is not to rule out most cases of belief, it will have to draw on a wider notion of truth-directedness, encompassing non-agential mechanisms that track the truth. Being regulated by any truth-tracking mechanism will then be sufficient for a cognition to count as truth-directed in the sense required for belief. (p. 499)

The argument they then present focuses on these “non-agential mechanisms”. It is these mechanisms that allow beliefs the formation of which is not accompanied by a conscious intention to believe truly to count as truth-aiming. How truth-tracking do these mechanisms have to be for their resulting mental states to count as beliefs? Can occasional non-evidential elements play a causal role in the formation of a belief, or are these doxastic states excluded by the definition because the cognitive process that produced them is not sufficiently truth-tracking? In order to answer this question, Shah and Velleman think the transparency of doxastic deliberation makes for a crucial difficulty for the teleologist. Because in order to obtain the result that every instance of asking oneself “should I believe p?” is transparent to the question “whether p?”, and is hence answered solely through what one perceives as evidence for the truth of p, the teleologist will have to stipulate that the truth-tracking of our belief formation is very strong. This kind of stringent truth-tracking however, would mean that the teleologist

would have to exclude many plausible candidate-beliefs from counting as beliefs. Wishful thinking is one such plausible candidate.

Being regulated for truth consists partly in responsiveness to evidence. The belief that *p* tends to be formed in response to evidence of *p*'s truth, to be reinforced by additional evidence of it, and to be extinguished by evidence against it. A crucial question is how strong these dispositions must be in order for the attitudes possessing them to qualify as beliefs. If belief can be influenced by evidentially irrelevant processes such as wishful thinking, then its responsiveness to evidence must be weak enough to leave room for such additional influences. And if this interpretation of truth-regulation is correct (as we believe), then the manner in which belief is regulated for truth can't be cited to explain the role of truth in doxastic deliberation. For when one deliberates whether to believe that *p*, this question not only gives way to the question whether *p* but does so to the exclusion of any other, competing question, such as whether *p* would be in one's interest. (p. 500-501)

Therefore, the teleologist is caught in a "dilemma": if she defines the sub-intentional processes that lead to belief as strongly truth-conducive, cases of wishful thinking and other states that intuitively seem like beliefs are excluded, but if the belief-formation processes are defined as only weakly truth-conducive, they cannot explain transparency.

#### §4.1.2 Normativism and transparency

Having decided that teleologism is not a sustainable option, Shah and Velleman offer their own account of transparency. Recall that on their view, one only counts as possessing the concept of belief if one accepts the norm "Believing that *p* is correct if and only if *p*". Doxastic deliberation, they argue, involves employing the concept of belief, since one asks oneself "should I believe that *p*?". In making use of this normatively constituted concept, subjects are automatically bound by the correctness norm, and hence cannot come to a belief outcome that differs from their answer to the "whether *p*?" question.

Here is the reason why considering whether *p* is the only means of deliberating whether to believe tout court that *p*. To deliberate whether to believe that *p* is to engage in reasoning aimed at issuing or not issuing in a belief that *p* in accordance with norms for such a belief. And to conceive of the attitude at issue in one's reasoning as a belief that *p* is to conceive of it as an acceptance that will be correct if and only if *p* is true. Thus, deliberating whether to believe must consist in reasoning aimed at issuing or not issuing in an acceptance that would be correct if and only if true, in accordance with norms for such an attitude. Since the norm of truth is already applied to the envisioned attitude as part of the concept of belief, it must be the norm in accordance with which deliberation aims to issue in the attitude. Aiming to arrive at an attitude in accordance with a competing norm would be incompatible with subsuming it under the norm of truth in conceiving of it as a belief. Deliberation whether to believe that *p* must therefore consist in reasoning aimed at issuing in an acceptance of *p* if and only if this acceptance would be correct in virtue of *p*'s being true. And the only way of reasoning with that aim is to consider whether *p*, so as to arrive, if *p* is true, at a judgment that would induce an acceptance of *p*. (p. 519)

The conceptual normativist way of conceiving of doxastic deliberation has obvious advantages over the teleological account. Because the norm of belief only plays an active role in belief formation in those contexts where the concept of belief is salient (as in the question "should I believe *p*?"), the normativist account can allow for differences in evidential sensitivity between contexts of doxastic deliberation, where belief formation displays perfect evidential sensitivity, and other contexts, where belief formation might be guided by non-evidential considerations as well, as is the case in wishful thinking. That nuance immediately protects the normativist from the dilemma suffered by the teleologist.<sup>40</sup> And there might indeed be a sense in which the more deliberate a subject is in the formation of her beliefs, the less possible it is for her to use non-

---

<sup>40</sup> It's a question I won't seek to answer here whether the teleologist could not make a similar move on her own terms that would free her from the dilemma: why not stipulate that belief's aiming at truth comes in degrees, and that it's only in the very deliberate contexts of formation that belief manifests perfect truth-directedness, whereas contexts where the formation of a belief flies as it were under the radar of the deliberating subject, its inherent aim is either less strong or more diverse?

evidential reasons to determine her belief. There seems to be a kind of wishful thinking that consists of a looking away from the evidence. The feat of this kind of self-deception has a lot to do with how explicitly one thinks about the matter at hand, and self-deceiving subjects often avoid the kinds of questions that would lead to doxastic deliberation. On the other hand, some self-deception is the product of laborious mental work, where the subject talks herself into a certain belief quite deliberately. Ultimately, the ratio of self-deceived beliefs between non-deliberate belief and belief that results from asking oneself whether one should believe will be an empirical question, which I suspect the nascent field of bias research may not have an answer to yet. On the philosophical level however, I think increased nuance in our predictions about how belief will function in different investigative contexts is a good thing. As we will shortly see, the picture that conceptual normativism rests on is still not quite nuanced enough when it comes to the evidential sensitivity of belief.

#### §4.2 Against the argument from transparency

I will evaluate the argument for conceptual normativism in three stages. Firstly, I will investigate the truth of the transparency claim about doxastic deliberation. Secondly, I will query in how far belief is unique in its transparency properties. Thirdly, I will look at Shah and Velleman's explanation of all these data.

##### §4.2.1 Is deliberation about whether to believe transparent?

Shah and Velleman qualify their claim about the transparency of doxastic deliberation substantially throughout their paper. Firstly, they contrast the question "whether to believe p?" with the question "whether I believe p?". The latter, they argue is not transparent, or at least not perfectly so. "Whether I believe p?" can express two different questions on their view: whether

one already believes  $p$ , or whether one from now on believe  $p$ . The former one can sometimes answer by asking oneself “whether  $p$ ?” if one manages to use that question solely as a “brute prompt” to cause an act of introspection, which will often deliver a correct answer to the question whether one already believes  $p$ , but not always. The question whether one from now on believes  $p$  can be answered through doxastic deliberation: one can make up one’s mind about  $p$  for the future by deliberating about  $p$ ’s truth, and hence obtain an answer to whether one from now on believes. That kind of transparency too, though, is imperfect. This is because one’s answer to the question “should I believe  $p$ ?”, and hence one’s decision about what to believe, doesn’t always succeed in actually determining what one believes. Shah and Velleman cite cases of irrational phobias as an example of such a failure:

Arriving at the judgment that  $p$  doesn't necessarily settle the question whether one now believes it, since one may find oneself as yet unconvinced by one's own judgment. One may reason one's way to the conclusion that one's plane is not going to crash, for example, and yet find oneself still believing that it will. (p. 505)

So even though the answer to the deliberative question whether one should believe is transparent, that answer does not always suffice to cause the relevant belief outcome, and what one actually believes is therefore not accessible to one just by asking and answering “whether  $p$ ?”.

But even the deliberative question “whether I should believe that  $p$ ?” itself is not always transparent to the question “whether  $p$ ?” according to Shah and Velleman’s account. This is because one can ask that deliberative question in a situation where one already has a cognitive attitude towards  $p$ , and the question now becomes whether to classify that attitude as a belief or not. To that classification question, the authors write, there is not always a clear answer:

whether I label my attitude toward  $p$  as a belief or as some weaker attitude can, in some cases, be up to me. In making that call, I need not be entirely guided by my answer to “whether  $p$ ?”, I can take some pragmatic considerations into account as well:

[O]ne can have pragmatic reasons for believing rather than suspecting that there is a predator in the shadows, and one can have pragmatic reasons for suspecting rather than believing that one has managed to prove Goldbach's conjecture. (As before, what we loosely describe as a choice between believing and suspecting is, strictly speaking, a choice between ways of classifying one's own cognition and, ultimately, a choice between norms implicit in the relevant classifications.) ... [T]hese are cases in which the deliberative question whether to believe that  $p$  is not transparent to whether  $p$ . (p. 517)

For that reason, not every instance of asking oneself “should I believe  $p$ ?” is transparent to one's answer to “whether  $p$ ?”. The deliberative question can be answered on pragmatic grounds “when asked in reference to an existing cognition classifiable either as a belief or as an hypothesis” (p. 518). The deliberative question *tout court* however, as Shah and Velleman put it, is strictly transparent. Now it's not entirely clear when doxastic deliberation is *tout court* and when it is qualified in the sense that does not guarantee transparency. The easiest way to evaluate the transparency claim then, is to only look at cases where there is no “existing cognition”: there transparency should hold without exception. Does it?

A first obvious failure of transparency is found in cases where one's answer to “whether  $p$ ?” is not “ $p$ ” or “ $\neg p$ ” but, for instance, “I don't know”. If I ask myself whether I should believe that my son is lying, where my stance on whether my son is lying is that I don't know, that clearly does not settle my question what I should believe: I still want to know whether I should believe. So let's stipulate that “I don't know” doesn't count as an answer to the question “whether  $p$ ” and reformulate the transparency principle as “One's answer to the question “whether I should

believe that p?" gives way to one's yes-or-no answer to the question "whether p?". But even if we constrain the "whether p?" question in this binary way, problems remain. Take firstly the kind of case that Shah and Velleman consider themselves: irrational phobias about airplane safety.

Ex. 4.1.: Percy is on an airplane and looks extremely anxious. The passenger next to him says to Percy: "Don't worry, the plane isn't going to crash". Percy was actually anxious about something else, he had not considered the question whether the plane is going to crash yet. But now that he hears his co-passenger's testimony, he asks himself "Should I believe that the airplane isn't going to crash?". He weighs the evidence at his disposal by asking himself whether the airplane is going to crash, and comes to the decision that it is not going to. His anxiety having in the meantime encroached upon his attitudes about airplanes however, he still keeps asking himself "Should I believe it?", reasoning around in circles along the lines of "Are we going to crash? – No, we're not. – But it would be really terrible if we did. – Should I really believe we won't? – I don't know – But we're not going to crash - etc."

Shah and Velleman grant that in cases like Percy's, the subject does not always end up with the belief that matches their answer to "whether p?". But it's not at all obvious that the deliberative question should fare that much better. It's perfectly conceivable that the question whether he *should* believe the plane won't crash will remain active in his mind, fuelled as it is by his anxiety. Here is another case:

Ex. 4.2: Henry, after a stroke of bad financial luck, finds himself homeless. His brother and sister-in-law kindly take him in until he can manage to pay rent someplace of his own again. Henry gratefully tries to be as agreeable a houseguest as possible, but he is

still embarrassed by the situation: he's not used to relying on charity. Since moving in with his brother, Henry has gained all kinds of evidence about his brother's private life that he would otherwise have never had: he sees paperwork lying around that points to shady business activities, he hears his brother come home late at night accompanied by a woman other than his wife, the house sometimes smells of illegal substances. The goings-on suggested by this evidence are all things Henry morally disagrees with quite strongly, but not quite so strongly that he sees himself morally obligated to intervene. Henry ends up suspending belief about whether his brother has a clandestine side-job, whether he is having an affair, and whether he uses drugs.

Here we have a case where someone asking themselves whether they should believe something bad about their brother explicitly does not lead to the question whether that bad thing is true, because the latter question is something that Henry would rather not think about. Loyalty to his brother makes him answer the question "Should I believe that my brother uses drugs?" negatively, without answering the question whether this is the case. Ex. 4.1 and Ex. 4.2 are both counterexamples to transparency. In the first, a subject answers the "whether p?" question, but that fails to answer the question "whether I should believe that p?" for him. In the second, a subject answers the "whether I should believe?" question without answering, or indeed asking, the "whether p?" one. So the first explanandum of the argument from transparency, that deliberation about belief is transparent, fails. Let's turn to the second explanandum: that its properties in deliberation set belief apart from other affirmative attitudes.

#### §4.2.2 Is deliberative belief uniquely transparent?

Earlier we qualified that the “whether p?” question in the transparency principle has to be understood in a binary way, so as to exclude answers like “I don’t know”. Confronted with cases like Ex. 4.2, that demonstrate the psychological compatibility of doxastic deliberation in the presence of perceived evidence and suspension of belief, the normativist might wish to make a further refinement: what’s at hand in transparency, is the question “whether I should believe p or believe -p?”, to which “neither” is excluded as a possible answer. This would be quite a concession, because ordinarily when we think about doxastic deliberation, about trying to decide whether one should believe something, suspension of belief is a salient possibility, so a definition of doxastic deliberation that excludes it as an outcome is not very intuitive. But say we nevertheless requalify the transparency principle to: “The question “whether I should believe p or believe -p?” is transparent to one’s yes-or-no answer to the question “whether p?”.

But with the “should I believe p?” question disambiguated in this way, the supposed difference with the other affirmative attitudes vanishes. Because if I ask myself whether to guess that p or whether to guess that -p, and my answer to “p or not p” is “p”, I will answer the deliberative question positively: I should guess p instead of guess -p. It is less common to ask ourselves this deliberative question for attitudes like guessing, suspicion or assumption, but that does not change the fact that if I *were* to ask myself “whether to suspect that p, or suspect that -p” whilst perceiving evidence for p, the deliberative question would be determined by my answer to the factual question to the same extent as is the case for deliberation about belief.

It seems to be the case that we have more control over what we guess and assume and estimate than over what we believe: we can let pragmatic reasons factor into whether we adopt one of

the weaker attitudes far more easily than we can let these pragmatic reasons factor into our beliefs. But this difference is found on the level of the decision between adopting an attitude and not adopting it (which includes the possibility of agnosticism), not the decision between adopting an attitude and adopting the negation of that attitude. The transparency-defender has to make a choice: either she defines transparency in such a way (believing vs not believing) that it marks a difference with the other affirmative attitudes, but in that case transparency for doxastic deliberation is false, or she defines transparency in a way as to shield it from suspension cases (believing vs disbelieving), in which case there is no difference between doxastic deliberation and deliberation about whether to guess, assume, estimate etc. The second explanandum, that belief is unique in its behaviour in deliberation, is either false, or true at the cost of the first explanandum. Therefore, the argument for conceptual normativism that is meant to be an inference to the best explanation, offers an explanation of a phenomenon that does not exist. Conceptual normativism stands in need of further defence.

#### §4.2.3 Does the normativist account work?

A second line of undercutting argument against conceptual normativism focuses not on what the normativist seeks to explain, but on how she goes about explaining it. Recall that on the normativist picture, anyone who possesses the concept of belief understands that belief is governed by the correctness biconditional. Anyone who deliberates about whether to believe *p* makes use of the concept of belief, and therefore conceives of herself as bound by the correctness biconditional. This is meant to explain why the subject cannot decide to believe something that she has judged not to be the case. We can take issue with this explanation, for

reasons that have been noted in the literature. First, consider the following point McHugh (2013) makes:

The (normativist transparency) explanation thus appeals to the subject's acceptance of a norm in order to explain the subject's being motivated to satisfy that norm to the exclusion of any other consideration. It thus seems to assume that accepting a prescriptive norm guarantees that you will be motivated exclusively by considerations relevant to how to satisfy that norm. On the face of it, however, this isn't so. There is a difference between accepting that, in doing something, you ought to satisfy a certain condition, and being motivated exclusively by considerations relevant to that condition. So Shah's explanation seems to involve a non-sequitur. (p. 453)

As McHugh points out, there is a difference between accepting a norm as true and being guided by that norm in your behaviour. Reflection about concepts that are undeniably normative can illustrate that this is not merely a theoretical difference. Promising<sup>41</sup>, for example, is clearly a normative concept: one does not grasp the concept of promising unless one grasps the norm of promising "Be so that, if you promise  $\Psi$ ,  $\Psi$  will obtain"<sup>42</sup>. Indeed, it is plausible to suggest that one only counts as promising if one understands that one is bound by this norm. Nevertheless, it seems obviously possible to deliberate about whether to promise to  $\Psi$ , and take factors into account that are not conducive to one's fulfilling the norm. Indeed, one can even deliberate whether to promise  $\Psi$ , and make the answer to that question entirely depend on whether one will *not* fulfil the promise, because it is one's explicit intention to violate the norm of promising. Another, even clearer example, is the concept of murder. Murder seems

---

<sup>41</sup> Steglich-Petersen (2006) mentions promising in an objection to the transparency argument as well, but he uses it to make a point about the possibility to violate, not about the possibility of deliberating to violate.

<sup>42</sup> The norm of promising might well be stronger than this. It might i.a. involve an additional epistemic condition to not promise things one does not know one can fulfil, or it might involve a narrow-scope ought, or it might involve a stipulation that  $\Psi$ 's obtaining needs to be caused by the promising subject. But the above seems minimally required for promising.

like a good candidate for a normative concept, the normative component of which says that it is by definition wrong. The correctness condition of murder is the empty set. If I understand this about murder, it seems that I can nevertheless deliberate about whether to murder someone, and use reasons in that deliberation that are irrelevant to whether it would be morally permissible for me to murder them, since everything is irrelevant to whether it would be morally permissible to murder.

Grasping the conditions of a normatively constituted concept and acting in accordance with those normative conditions can indeed come apart. Subscribing to certain norms has some motivational force, but Shah and Velleman have not at all shown that this force is as immediate and irresistible as their normative explanation of doxastic transparency assumes. This means that even if we take transparency of doxastic deliberation, or some weaker principle that we might settle on, seriously as an explanandum for a theory of belief, stipulating that the concept of belief has normative content does not even suffice to offer that explanation: even with clear cases of normative concepts, there is no guarantee that the subject's grasping of the norm in question will play the kind of strong motivating role in deliberation about whether to perform the norm-governed act in question.

There is another problem with Shah and Velleman's story about doxastic deliberation. At the heart of their account is the idea that transparency happens in contexts of doxastic deliberation ("tout court"<sup>43</sup>), making for a strong kind of evidential sensitivity, even though in other contexts of belief formation, there can be causal reasons for belief that are not at all relevant to the truth of p:

---

<sup>43</sup> See the quote in §4.1.2.

This explanation of transparency leaves room for the possibility that beliefs can be influenced by non-evidential considerations, because it entails that one is forced to apply the standard of correctness only in situations in which one exercises the concept of belief. Not all belief-forming processes require the subject to deploy the concept, and the norm of truth that controls doxastic deliberation needn't control other processes. Our explanation of transparency thus allows for the fact that passions can influence belief. (p. 501)

And this difference between two kinds of belief formation (transparency in one case, possibility of wishful thinking in the other) is supposed to be due to the concept of belief making an appearance in the former.

[B]ecause the norm is contained in the concept of belief, and doxastic deliberation is framed by the question whether to *believe* that *p*, anyone engaging in doxastic deliberation knows that this is the relevant norm. (p. 503)

But do I really need to ask myself the question "should I *believe* that *p*?" in order to count as deliberating about whether to believe? Can't I engage in the same process without thinking explicitly about belief? If I just ask myself "is it really the case that *p*?" inquisitively and probingly and after weighing my evidence come to judge that *p* is the case which causes me to believe that *p*, isn't that belief a product of doxastic deliberation too, a kind of doxastic deliberation that we would expect to be just as transparent to the "whether *p*?" question as the one that explicitly mentions belief in the deliberative question? Curiously, Shah and Velleman seem aware of this possibility:

[D]eliberation whether to believe that *p* need not be initiated by an articulation of the deliberative question: one can start right in with the factual question whether *p* and yet be recognizably deliberating whether to believe. (p. 505)

But if it is not necessary for doxastic articulation to articulate a question that features the concept belief, it's far less obvious how the explanation of governance by the conceptual norm is

supposed to work. For this reason, too, the argument from transparency does not give us cause to accept conceptual normativism about belief.

### §4.3 Rebutting conceptual normativism

Thus far, I have argued that conceptual normativism fails to be established through the argument from transparency. I now turn to rebutting arguments against conceptual normativism that seek to discredit the view independently of how the normativist attempts to establish it. A first challenge takes the falsity of state normativism, which I have sought to demonstrate in Chapter 3, as a premise. Once we discover that the mental state of belief is not in fact intrinsically governed by a norm, we might see that as a reason to revise our concept of that mental state. If the conceptual normativist wants to stick to a norm intrinsic to the concept, this discrepancy will need some argument to be made plausible. In what follows, I focus on objections that are independent from those against state normativism.

#### §4.3.1 Conceptual disagreement

The first general problem with conceptual normativism I want to discuss stems from reflections about conceptual disagreement. Recall that Shah and Velleman think the correctness biconditional applies to belief as a matter of conceptual truth, and that anyone who reasons about belief is aware of it. They are thus committed to an epistemic conception of conceptual truth, such that competence with the concept of belief ensures acceptance of the correctness biconditional. Williamson (2006, 2007) gives reason to think there are no epistemic conceptual truths, because for any candidate for such a truth, there could be competent users of the relevant concepts who fail to be disposed to assent to it.

Williamson's argument draws on the possibility for disagreement about the truth of some proposition like "every vixen is a fox" due to the background theoretical beliefs of competent concept users. For the candidate conceptual truth that belief is correct iff true, such disagreement is real and by no means marginal. Although the correctness biconditional that Shah and Velleman propose as part of the concept of belief is relatively popular in the contemporary literature on epistemic norms, disposition to assent to it is nowhere near as widespread as to the conceptual truths Williamson argues to be non-epistemic. There are those who think the norm of belief has a different criterion than truth (Littlejohn, 2013; Williamson, 2002; Smithies, 2012b) and there are even some that reject the notion that belief has any intrinsic norm whatsoever, such as Papineau (2013) and myself. So even if Williamson is wrong and epistemic conceptual truths do exist, the correctness biconditional is looking like an unlikely candidate. If it is not a conceptual truth in the epistemic sense (so that all who reason with the concept of belief are disposed to assent to the correctness biconditional), Shah and Velleman will need a different explanation for its supposed force over belief formation. Its being a conceptual truth is made unlikely by the philosophical disagreement because the disagreement would have to be interpreted as a sign that the dissenting parties fail to grasp the concept of belief.

This is implausible firstly for epistemic reasons. On Shah and Velleman's account, understanding one's beliefs to be subject to the correctness biconditional is automatic. No reflection is required of the subject to come to this insight; knowledge of belief's correctness condition is immediate. How is it possible then that so many philosophers would have gotten things wrong? The conceptual normativist should provide some explanation of the existence of serious

philosophical disagreement about the norm of belief, given how easily knowledge of that norm is supposed to come on her picture.

Excluding people who disagree about the norm of belief from grasping the concept of belief furthermore has the deeply problematic consequence that these dissenting philosophers not only fail to see something that is meant to be obvious, but that they are not even talking about the same thing as the conceptual normativist: there can be no disagreement about the norm of belief, since anyone who doesn't accept the norm is by definition not talking about belief but about something else. This is a problem for the project of philosophy as concept analysis generally. But for conceptual normativism about belief, things are even worse, for this view not only ties the possession of the concept of belief to grasping its norm, it also ties the possession of the concept to the capacity for doxastic deliberation. In other words, on the conceptual normativist view, philosophers like Littlejohn, Williamson, Papineau and myself could have unreflective beliefs that arose from sub-intentional cognitive processes that they have never consciously conceived of as beliefs, but they cannot hold deliberate beliefs. All of their convictions held as a consequence of deliberation fail to be beliefs. This will exclude them from believing the majority of their philosophical views, but also extends much further into their everyday convictions. I take it that this consequence of disagreement about the norm of belief is a considerable problem for conceptual normativism.

#### §4.3.2 Lawless deviance

With the above in place, we can see that a version of the lawless deviance objection applies to conceptual normativism as well. Here's a case:

Ex. 4.3: Lucinda doesn't care much about epistemic norms. She accepts that it's better for her beliefs to be true about things like which food is poisonous, what time the movie starts, and who her boss is at work but for anything that has little practical bearing on Lucinda's life, she rejects the notion that her beliefs ought to be true. So Lucinda goes through life picking up beliefs without much scrutiny, such as that hand wrinkles predict one's future (she was told this by a hand reader advertising her services in the shopping mall) and that AIDS was caused by bestiality (she overheard a 10 year old boy asserting this to his friends). When she encounters evidence that these careless beliefs might be false, the kind of evidence that would usually prompt one to lower confidence or re-investigate the matter, Lucinda does nothing and continues believing, not because she does not grasp the defeating evidence but because she explicitly reasons that it's not important to her whether these things are true and that having false beliefs about them is fine by her.

Lucinda is some kind of epistemic sociopath. She is aware that many people accept a certain norm and guide their behaviour in response to it, but she does not accept this norm as applying to herself. Presumably, Lucinda is somewhat differently disposed from most people on the level of her mental states, but most importantly there is a conceptual difference: she disagrees with the truth norm of belief, and some of her false beliefs are false because of this disagreement. Now take one of these beliefs, which is false, formed in response to flimsy evidence, and sustained in the face of defeating evidence *because* Lucinda decided it's not important to her that the belief is true, and applies no norms to her beliefs other than ones instrumental to what is important to her. This kind of deviant belief will fail to count as a belief according to Shah and Velleman's theory. Interestingly, they themselves mention the importance of agents' self-

conceptualisation of being governed by the truth norm to the question whether they count as believing:

[I]mplicit in the concept of belief, as we understand it, is a policy of deference that goes beyond expressing respect for the subject's rational agency. According to our analysis of the concept, one may apply the norm of truth to a cognition by classifying it as a belief only if it is in fact regulated for truth to some extent, as if it were being held to the same norm by the subject himself...The concept of belief, as we understand it, doesn't require one to assume that the subject is generally correct; it merely forbids one from holding him to a standard of correctness that he has no disposition to meet. (p. 516)

Shah and Velleman consider the possibility of cases where we cannot hold a subject to epistemic censure because she does not qualify as a believer due to the normative criterion. In other words, exactly the kind of scenario the lawless deviance objection focuses on. But Shah and Velleman don't consider deviant subjects, epistemic sociopaths, chronically lazy inquirers, subjects who convince themselves some cognitions of theirs are "just assumptions" that they don't need to double check even though they're acting and reasoning from them as if they were beliefs. They focus on subjects who carefully deliberate about whether to classify a cognition of theirs as a belief or a hypothesis based on their willingness to put the proposition at risk of disconfirmation: "Because one can often decide to go on testing an accepted proposition for truth, however, one can justifiably decide to conceive of it as an hypothesis rather than a belief." (p. 513) But what if one does so non-justifiably? What if one operates by much more lax norms of what a belief is, like Lucinda does, or what if one is deceiving oneself about one's true intentions regarding further testing of one's "hypothesis"? Should we let those subjects go without epistemic censure? We should not. Epistemic norms cannot be opted out of, and they certainly cannot be opted out of *in virtue of* epistemic deviance. Conceptual normativism entails the possibility of such opting out through conceptual disagreement with

the norm or through self-deceived misapplication of the norm, and hence it entails lawless deviance.

### §4.3.3 State-differentiation

Like lawless deviance, the next problem for conceptual normativism is also shared by several of its normativist sister theories: the problem of mental state-differentiation. Shah and Velleman bring up the need for state-differentiation in order to motivate the need for their work on transparency: "Our question then becomes how accepting a proposition in the manner of belief differs from other modes of acceptance." (p. 498). The answer to that question, on their view, is twofold: belief is regulated for truth and belief is governed by the standard of correctness. These two factors in conjunction are meant to explain why belief functions differently from other affirmative attitudes, where transparency is one of the key functional differences. As I explained in §4.2.2, this difference is misleading: when doxastic transparency is formulated in a way immune to counterexamples, the other affirmative attitudes are just as transparent. Furthermore, I have argued in Chapter 2 that there is a general problem for state-differentiation through correctness conditions, because of the dilemma of affirmative state correctness.

Recall the dilemma: in ordinary language, we ascribe correctness to other affirmative states under the same conditions as those under which we ascribe it to belief. Usually, although not always, that condition is truth. So the normativist has two options. Firstly, to argue that belief is normatively equivalent to the other affirmative states: it is permissible (obligatory, forbidden, good, bad, fitting) in the exact same circumstances. This normative indistinguishability is a highly

unintuitive result.<sup>44</sup> States like assumption or estimation seem like exactly the kinds of attitudes one ought to adopt in situations where belief is not permissibly available. Secondly, the normativist can argue that even though we can speak of assumptions and estimations as correct in the same circumstances in which we speak of correct belief, the correctness biconditional as a normative principle only follows for belief. But since the plausibility of the correctness biconditional for belief is rooted in ordinary language practices, this move is not just *ad hoc* but actively undermines the case for belief's presumed norm in the first place. This is the affirmative state dilemma as applied to the argument from correctness for the truth norm of belief. As such, it is meant to establish that belief's correctness conditions cannot be used as evidence for the norm of belief. The dilemma therefore has bearing on the content of the belief norm. But it can be brought to weigh against normativism generally as well.

#### §4.3.4 Psycho-ontological indistinguishability

A version of conceptual normativism that would argue that the correctness standard is a sufficient condition for belief, would immediately fall prey to the following dilemma. Beliefs are said to be correct in the same circumstances in which the other affirmative states are said to be correct. Therefore, the first horn of the dilemma becomes that beliefs are not just normatively equivalent to the other affirmative states (normative indistinguishability), but generally identical to all the other affirmative states that share its correctness conditions (psycho-ontological indistinguishability<sup>45</sup>). But conceptual normativism does not entail that the correctness condition

---

<sup>44</sup> In addition to being unintuitive, normative indistinguishability also leads to the problem of overprescription that we will discuss in Chapter 7.

<sup>45</sup> Psycho-ontological indistinguishability is a worse result for conceptual normativism than for essential normativism. If it were true on the essential level, that would mean that all affirmative states are in fact

has to be our criterion of state-differentiation. The normativist may be wrong about the uniqueness of the correctness condition to belief, but as long as she stipulates it to be a merely necessary condition, some other criterion might still help avoid the psycho-ontological indistinguishability of affirmative attitudes (although this move does not protect normativism against the normative indistinguishability horn of the original dilemma). And such a multi-criterion definition of belief seems to be in the spirit of several normativist accounts, so that one might hope that the oversight of affirmative attitudes' shared correctness condition will be without this radical consequence. Take the following quote from Lynch (2009):

Three considerations suggest [the correctness condition] is, moreover, a constitutive or essential fact about belief. First, the fact that truth is a norm of correctness for believing *is part of what* distinguishes believing from other cognitive attitudes. Imagining, assuming, and hoping, for example, are each governed by norms – assumptions can be justified or not, imaginings can be sharp or vague, hopes can be rational or irrational. But neither imagining that p, assuming that p, nor hoping that p are properly evaluated in terms of truth. Believing is. (p. 229, italics mine)

Of course, transferring the responsibility of state-differentiation to some other necessary condition undercuts the motivation for normativism somewhat. Many normativists have argued, like Lynch above, that the (presumed) uniqueness of the norm that governs belief supports the intrinsic status of that norm. Wedgwood (2002) for instance writes:

There are reasons for thinking that these epistemic norms articulate essential features of the types of beliefs that they apply to ... [I] seems plausible that, quite generally, types

---

of the same kind: there is no difference between guesses, assumptions, suspicions, beliefs, hypotheses etc. qua mental states. That result would be highly surprising, but surprising discoveries about the nature of our minds are possible given our limited knowledge at present. If psycho-ontological indistinguishability were true about our concepts of mental states like guessing, hypothesising, assuming, suspecting, and believing, this would entail widespread and systematic error in the way we have thus far, both in our folk theory of mind as well as in epistemology, reasoned about affirmative attitudes. That there is no conceptual difference between belief and what we think are different kinds of affirmative attitudes is highly unlikely.

of mental state are individuated by the conditions under which they satisfy normative concepts. (p. 270)

Shah and Velleman place less explicit motivating weight on state-differentiation than Lynch. Their argument for normativism relies mostly on its explanation of transparency. But note that that argument in turn rests on the power of the correctness condition in conjunction with belief's evidential sensitivity to differentiate belief from other cognitive states, to explain why transparency is unique to belief. As we have seen, it is far from obvious that transparency is unique to belief. Could the two properties of belief that were meant to be at the root of it nevertheless suffice to individuate belief conclusively? That correctness does not perform this differentiating task we have concluded. Shah and Velleman's other sufficient condition for belief, its being regulated for truth, is not a normative one. A definition of belief that relies on this descriptive condition to differentiate belief from other affirmative attitudes is no longer normativist narrowly construed: the norm of belief is entailed, but does not need to be mentioned in its definition, for it will be the norm of all affirmative attitudes. Without the norm as concept-differentiating, doxastic normativism becomes a weaker theory, and problematically it will still have to account for normative indistinguishability. For that reason, before investigating state-differentiation through Shah and Velleman's descriptive criterion in §4.4, I turn to some possible alternative routes for doxastic normativism to use a normative property to define and differentiate belief.

Most normativists we encounter throughout this thesis favour a truth norm of belief as its normative element. The correctness condition features in many of these accounts, either as the formulation of the truth norm itself (Engel, 2013a; Shah & Velleman, 2005; Wedgwood, 2002), or as the fact about belief that grounds the norm (Engel, 2001; Gibbard, 2005; McHugh, 2012b;

Raleigh, 2013; Whiting, 2010). But in my definition of doxastic normativism, both broadly as well as narrowly construed, I did not mention correctness. Any kind of normative property included in or entailed by the definition of belief makes that definition normativist. Given that correctness turns out not to be belief-defining after all, maybe a different normative property would be better.

#### §4.3.5 Normativism without correctness

It is a limitation imposed by the scope of this thesis that we cannot investigate in detail the implications for normativism of different contents of the belief norm: truth norms, knowledge norms, obliging norms, restrictive norms, evaluative norms, instrumental norms etc. The conclusion features a summary of arguments presented mentioning which formulations of normativism they work against, but not all of those classifications can be explored in detail. In this section I offer some relatively brief observations for the feasibility of normativist projects that use a truth or knowledge norm without appealing to correctness conditions.

*Truth-norm-normativism.* Affirmative attitudes are all truth-apt: the truth conditions of their propositional contents provide truth conditions of the attitude towards that proposition, in contrast to non-affirmative propositions like fear or hope that cannot be said to be true in the same way.<sup>46</sup> For normative predicates other than “correct”, it is certainly not the case that their normative relationship to truth is identical: beliefs and guesses are correct in the same circumstances, but there are situations where I ought to believe p rather than guess p. But it will be difficult to capture that difference that makes for normative distinguishability if the norm

---

<sup>46</sup> When we speak of “true hopes”, this either means “a hope I truly have” or “a hope that came true”. It does not mean a hope with a true propositional content at the time of ascription (although it is of course compatible with the proposition being true).

of belief has truth as its criterion. For all affirmative attitudes are such that they are *ceteris paribus* better when they are true than when they are false. It might be more permissible for a guess to be false than for a belief to be false, but it nevertheless would have been even better by the lights of epistemic normativity if the guess had been true. And that translates into difficulty for the normative predicate one uses in one's definition of belief.

Consider first the difficulty of formulating obliging belief norms: if it is obligatory for some subject to believe *p* because *p* is true, the normativist needs to make sure that it is not also obligatory for that subject to guess, suspect, hypothesise, estimate, and conjecture that *p*, because that would oblige the subject to something psychologically impossible. But it's not obvious how to justify this kind of difference in obligatoriness of holding attitudes with true content between the different affirmative attitudes on a truth norm. Beliefs and guesses etc. don't differ in their relationship to truth: they are made true by the same things, for all of them it is usually better for the subject if their content is true rather than false. Yet the normativist needs to offer some story as to what makes belief different from the other affirmative attitudes so that true beliefs about some proposition can get obligatory status while true guesses about the same proposition do not, on pain of generating a truth norm for all affirmative attitudes that will demand the psychologically impossible. This is the overprescription risk of obliging truth norms.

Consider next deontic norms that involve only restriction and permission (believing *p* is permissible only if *p*). There the overprescription problem does not arise. But instead, there is an underprescription problem: without obliging norms, a subject who systematically avoids epistemic responsibility by suspending belief does much better by the lights of the epistemic

norm than one who takes risks in trying to extend her knowledge, as does the subject who systematically avoids difficult areas of investigation, settling only on safe affirmative attitudes in order to avoid censure by the belief norm. Pyrrhonian scepticism might not be a psychologically feasible policy, but even the hypothetical Pyrrhonian should not get hailed as the ideal epistemic subject, put on the same evaluative footing as the omniscient subject. We need obliging epistemic norms, and the truth norm has difficulty allowing for that.

Consider next evidential norms, that make epistemic justification the criterion of the belief norm. Again, it is not clear how to differentiate between beliefs and other affirmative attitudes in this respect: all of them can be based on varying degrees of evidence, where the same things count as evidence<sup>47</sup>, and guesses, assumptions, suspicions, estimations etc. that are based on more evidence are again said to be better, more rational, more normatively justified than ones based on less evidence, to the same extent that beliefs with more evidential support are said to be better in this way. The dilemma of over- and underprescription arises again.

At this point, the normativist might want to help herself to the distinction of *normative force*. Beliefs might be like guesses in the sense that they are better when they are true, but they are unlike guesses in the sense that it is more important for them to be true than it is for guesses. The normative distance between false and true beliefs is bigger than that between false and true guesses, assumptions, estimations etc., and the same holds for evidence. And that difference in normative distance seems like exactly the reason why, when faced with the risk of a mistake, it is better to adopt one of the weaker affirmative attitudes. Holding a false belief is

---

<sup>47</sup> Facts, perceived facts, knowledge,...whatever one's preferred conception of evidence is, those things that are evidence for a belief are also evidence for a guess, hypothesis, assumption etc.

much worse than holding a true belief, while holding a false weak affirmative attitude is comparatively less worse than holding a true one, which favours trying one's luck in contrast to belief. I think this is exactly the right way to think about the normativity of affirmative states, but it won't provide the normativist with a solution to state-differentiation. Consider the kinds of norms that have been proposed as intrinsic to belief:

For any  $S$ ,  $p$ :  $S$  ought to believe  $p$  only if  $p$  is true

$S$  may believe  $p$  iff  $p$  is true

$S$ 's belief that  $p$  is good iff  $p$  is true

None of these are suitable to capture a normative difference between affirmative states, as they will all apply to all of them. Take the first of them, that forbids false belief. It is also the case that I ought not to guess  $p$  when  $p$  is false, ought not to assume  $p$  when  $p$  is false, ought not to suspect  $p$  when  $p$  is false,...for if I do so, I will have made a mistake. The difference between belief and the weak affirmative attitudes is not that only belief ought to be held only if true, but that the force of this ought is somehow stronger, which makes violations of the belief ought worse than violations of the others. The force difference is captured in a comparative judgment, but comparative judgments cannot be accommodated on doxastic normativism. It cannot be part of the concept of belief that it is more important for it to not be false than for other guesses and estimations, because that would involve reference to other mental states in the concept of belief, even though belief seems clearly metaphysically independent from these other mental states: there is a possible world where beliefs exist but not guesses and estimations, so a normative comparison to guesses and estimations cannot be part of the concept belief.

The normativist might try to offer something like the comparative judgment in her definition of belief by introducing graded oughts to the various norms. I don't think this would be a very intuitive way of talking about epistemic normativity, and it would presumably be very difficult to spell out formally, but it won't do any good for state-differentiation anyway. I'm only able to offer a brief sketch for why graded oughts won't work. Take the restrictive deontic truth norm: "S ought to believe p only if p is true". For ease of formulation, consider this norm under its violation form: S is forbidden from believing p falsely. Now combine this with the idea of "forbidden-n", where n is supposed to capture how bad the infraction is. So say for the sake of argument that false belief that p is forbidden-10, and false guess that p is forbidden-5 and false estimation that p is forbidden-7. Clearly, if this were the case, subjects would have a normative incentive to adopt as weak an affirmative attitude as feasible towards all propositions under consideration, for if they get things wrong, they will be in less significant violation: underprescription again. To avoid the consequence that subjects ought to hedge their doxastic bets this way, we need obliging oughts as well. Supplement the graded prohibitions then with the following graded injunctions: true belief that p is obligatory-10, true guess that p is obligatory-5, true estimation that p is obligatory-7. But that cannot avoid the fatal consequence that the subject must believe and guess and estimate that p, which is impossible since these are incompatible mental states<sup>48</sup>. We're back at the problem of overprescription of normativist truth norms.

*Knowledge-norm normativism:* What all of this suggests is that we may want to look for a criterion for the belief norm that belief stands in a different normative relationship to

---

<sup>48</sup> See §7.1.3 for discussion.

compared to the other affirmative states. Knowledge immediately comes to mind. On a belief norm like "Believe only propositions you thereby come to know", we do not get the troublesome consequence that the same injunction holds for weak affirmative attitudes, because we cannot come to know things by merely guessing or assuming them: we have to believe them. The difficulty for a normativist definition of belief using a knowledge norm is that belief and knowledge might be asymmetrically conceptually dependent: the concept of knowledge clearly involves the concept of belief, but it is not clear that the concept of belief involves the concept of knowledge. It might be possible for belief to exist in a world where knowledge does not exist, and if that's the case, doxastic normativism cannot take on a knowledge criterion, because that would include reference to knowledge in the definition of belief.

Asymmetrical conceptual dependence seems like a less serious problem for a definition of belief than normative and psycho-ontological indistinguishability and over- and underprescription. The knowledge-norm normativist might simply say that our intuition about the possibility of belief existing in a world without knowledge is mistaken. Knowledge-norm normativism would however still face the lawless deviance objection. A final objection to doxastic normativism, the objection of state-overarching normativity, is discussed in Chapter 7. That objection will hold against knowledge-norm normativism as well.

To sum up, state-differentiation related problems bear against all versions of normativism that are deontic and take truth as the criterion. Evaluative truth-norms fare somewhat better on this front, but share other difficulties with the deontic forms. Knowledge norms do significantly better on state-differentiation.

#### §4.4 Non-normative state-differentiation

The normativist, especially one who is inclined towards truth norms, may now want to attempt shifting the burden of state-differentiation from the belief norm to some other, descriptive, element of belief. As mentioned, this would take some of the motivation for a normative definition of belief away, and it wouldn't solve the normative indistinguishability problem for the truth norm versions. It would be coherent though with the way we defined normativism that the norm essential to belief is not what distinguishes it from other mental states. I now turn to the feasibility of providing an alternative distinguishing criterion. Shah and Velleman think that belief has two necessary conditions: the correctness condition and its being regulated for truth. They think both of these are unique to belief. Are they right about the uniqueness<sup>49</sup> of truth regulation? Shah and Velleman don't elaborate much on their reasons, but on reflection it seems clear they cannot be right. There are obvious ways in which the weak affirmative attitudes are regulated for truth. Guessing and assuming show weak evidential sensitivity, where a lack of evidence either for or against the proposition is compatible with their formation, but evidence against the proposition does prevent the subject from adopting these attitudes: you cannot get yourself to assume that there is a pink elephant in the courtyard if you have evidence that there is no pink elephant in the courtyard (for instance, you're looking at the elephantless courtyard right now). Suspecting and estimating show stronger evidential sensitivity, where at least some positive perceived evidence is necessary for their formation: you cannot suspect that the bomb

---

<sup>49</sup> "[B]elief is regulated for truth, whereas other, non-belief-involving cognitive attitudes are not." (Shah & Velleman, 2005, p. 498).

is behind the third door to the left without any evidence to that effect whatsoever. Being regulated for truth is not unique to belief.

In light of the above, the normativist might try marking the difference between belief and the other affirmative states quantitatively: they are all evidentially sensitive to an extent, but belief more so than the others. That strategy won't work either though, because of the amount of psychological variability different people demonstrate when it comes to attitude formation. For belief, this has been well-studied in terms of need-for-closure levels, which differ from person to person and within one person from context to context<sup>50</sup>. It seems clear that the amount of evidential input required for the other affirmative attitudes is similarly variable: some people form suspicions on much less evidence than others, some people make even assumptions only very carefully. That means that stipulating an absolute threshold of evidential sensitivity to define belief will not be possible. Rather, the threshold is likely to be a comparative one: for any agent, regardless of how much evidence it takes for her to guess that *p*, and of how much evidence it takes for her to believe that *p*, the latter amount is higher. But as above for definitions invoking normative force, comparative mention of other mental states is problematic when trying to capture the essence of belief.

The crucial difference in terms of evidential sensitivity between belief and the weaker affirmative states seems to be a normative one rather than one of psychological law: belief is the kind of state that *ought* to take more evidence as its input condition. But trying to capture that

---

<sup>50</sup> Need-for-closure captures the speed with which subjects tend to settle (reach doxastic closure) on a belief as opposed to continuing their investigation (Kruglanski, Webster, & Klem, 1993). The need for closure differs from person to person (Webster & Kruglanski, 1994) but within one person it can also differ from situation to situation (Van Hiel & Mervielde, 2003).

normative property is exactly the challenge we encountered above. If not its connection to truth, its correctness condition, or its psychological relationship to evidence, what is it that makes belief such that it ought to display more evidential sensitivity than other affirmative states and simultaneously that it is subject to a stronger kind of epistemic norm than the weaker affirmative states? One might try to appeal to belief's role in causing action.

#### §4.4.1 Belief and action

It is a well-known feature of beliefs that we take their contents as premises in our practical reasoning. I deliberate whether to try to extinguish the fire or whether to take the fire exit immediately, the premise that there is a fire is taken as given in this process. Whichever action I end up performing would be inexplicable without my belief that there is a fire. Had I never performed either action, for instance because someone else extinguished the fire before I made up my mind, my belief would not have been acted upon but would nevertheless have disposed me to act on it. Had I not deliberated about what to do and instead ran towards the fire exit immediately, the belief that there is a fire would nevertheless have played a crucial causal and explanatory role in my behaviour. Therefore, it makes sense to express the relationship between belief and action dispositionally. What we believe tends to be used as a premise in our practical reasoning. Were we to reason about something to which the content of our belief is relevant, and the action resulting from that reasoning will be as if the belief were true. And with or without practical reasoning, when we believe  $p$ , if  $p$  were to become such that we can act as if it were true or act as if it were false, the belief disposes us to act as if it were true. If I believe the fire exit is on the right, I am disposed to act as if it is on the right. The term *acting on  $p$*  will be used to denote acting as if  $p$  were true, which can happen as a result of practical reasoning that

features  $p$  as a premiss, or more immediately. Call the following the strong belief-action principle:

Def. 4.1 Strong Belief-Action: For any  $S$ ,  $p$ :  $S$  believes  $p$  if and only if  $S$  is disposed to use  $p$  as a premise in practical reasoning and if  $S$  is disposed to act on  $p$ .

Strong Belief-Action is a version of a dispositionalist definition of belief, of the kind that goes back to philosophers like Ryle (1949) and Braithwaite (1932). This is a different kind of dispositionalism than the one we encountered in Chapter 2, where belief was defined in terms of the dispositions that *lead to* belief. The dispositionalism of Belief-Action is one of defining belief in terms of what it disposes one to do: it is about the effects of belief, not about its causes. A dispositional definition of belief may well have to be more expansive than Belief-Action: contemporary dispositionalist Schwitzgebel (2002) for instance argues that phenomenal aspects need to be included along with behavioural ones in the set of events belief is disposed to cause. The reason I am focusing on action is that action, as opposed to speech acts or phenomenal experience, seems to have the best chance of being *unique* to belief.

With these clarifications in place, let us look at how the doxastic normativist might try to use a principle like Strong Belief-Action. She could argue as follows: belief is different from the other affirmative states. This difference is one in disposing towards action: belief disposes the subject to act upon the proposition believed, the other states do not.<sup>51</sup> Since acting on falsehoods carries negative consequences, this makes it more important for beliefs that they are epistemically in good shape, than for the other affirmative attitudes. The tendency to give rise

---

<sup>51</sup> Shah and Velleman don't actually think that action-disposing is unique to belief, they think that one can act on hypotheses and the likes of it as well. So this is a construction of what a normativist in their position might wish to argue given the failure of other state-differentiating criteria discussed above.

to action is intrinsic to belief. Therefore, belief is subject to an intrinsic epistemic norm. Call this the *simple argument from action-disposing*.

There are two ways to attack the simple argument. Firstly, by showing that action-disposing is not as unique or intrinsic to belief as suggested. Secondly, by focusing on the fact that the weaker affirmative attitudes are still subject to *some* normativity, even though the norms governing them seem less important in some sense or other. If the norm of belief is entirely grounded in its action-inspiring potential, this limited norm-governedness of guesses, suspicions, assumptions etc. would be entirely mysterious. I turn to exploring strong Belief-Action in §4.4.2, and then to the causal role of the weaker states in §.4.4.3

#### §4.4.2 Exceptions to strong belief-action

Much discussion exists about the left-to-right direction of principles like Strong Belief-Action. The classification of delusions in particular is hotly debated: since some delusions do not seem to dispose subjects in ways we would expect of a belief, some dispositional definitions will rule these delusions to be cases of non-belief (Marcus' (1990) account for instance has explicitly been used in an argument to this effect (Wilkinson, 2013)). Some pro-dispositionalist voices include Tumulty (2011), Currie (2000), and Schwitzgebel (2002), for arguments against action-disposing being a necessary condition on belief see Bayne and Pacherie (2005), Miyazono and Bortolotti (2014), and Rose, Buckwalter, and Turri (2014). For our current purposes, the right-to-left direction of Belief-Action is more important. Can one only act on p if one believes p? There is some existing debate about whether a certain motivational role suffices to characterise something as a belief. Authors have drawn on alief-like cases (Gendler, 2008, Schwitzgebel, 2010), the motivational profile of imagination (Velleman 2000, Van Leeuwen 2009), and cases

of absent-minded behaviour (Zimmerman, 2007). In this section, I am specifically interested in the question whether action-disposing is sufficient to distinguish belief from other affirmative attitudes.

Can we not act on suspicions or guesses, treat assumptions as premises in our practical reasoning? This matter is a difficult one to settle, and ordinary language might be misleading here to an extent. Firstly, consider the expression "acting on a suspicion". It is very natural to speak of acting on suspicions, for instance in this kind of context:

Ex. 4.4: I tell my best friend I suspect my husband of cheating, and, trying to ascertain how serious I am, she asks "Have you acted on this suspicion?", to which I respond "Yes, I called a private detective yesterday."

But now contrast the use of "acting on" in Ex. 4.4. with "acting on a belief" above. I am not actually acting on the proposition "my husband is cheating": acting on in the relevant sense means taking as given in one's actions, and I am clearly not taking that proposition as given, but as to be established. Indeed, if I believed that my husband was cheating, I probably wouldn't be calling a private detective, but maybe a divorce lawyer instead. So "acting on p" is ambiguous between taking action as a consequence of holding some attitude that p (even non-affirmative attitudes, like fears, can be said to be acted on in that sense) on the one hand and acting as if p were true on the other. It's only the latter kind that is at issue in Belief-Action. Here is a case though where the behaviour referred to by "acting on" is behaviour as if p is true:

Ex. 4.5: I'm standing at a forked road with Fannin Street on my left and Jubilee Street on my right on my way to the local library, and I know that the library is on either Fannin or Jubilee but can't remember which one. Jubilee Street looks from where I'm standing

like it's a little more densely built whereas Fannin looks to be more suburban, so I guess that it's Jubilee Street and proceed to the right. When ten minutes later I reach the library, I say to the librarian letting me in "I'm glad I found it, I was just going on a guess about which street you were on".

It sounds a lot more plausible to describe my action here as legitimately acting on my affirmative attitude than in the private detective case. However, a relevant feature of Ex. 4.5 is that if I had been wrong, I could have gone back to walk down Fannin Street instead.<sup>52</sup> Paradigm cases of acting on a belief on the other hand are much less sensitive to stakes. Nevertheless, Ex. 4.5 could well be an example of acting on a guess. Now take a straightforward case of assumption:

Ex. 4.6: I have to schedule a group meeting and am not able to check the other members' availability before booking the meeting room. I know most of their schedules relatively well but Nick has just joined our group the day before and I know nothing about him other than that he is 35. I reason as follows: I assume Nick has children, who would be primary school age, so he's more likely to be free during school hours, so I'll schedule the meeting for 2pm.

In Ex. 4.6 it seems accurate to describe me as acting on the affirmative attitude in question: I use the assumption that Nick has children as a premise in my practical reasoning, I act as if it were the case, I treat it as given. However, in a case where the stakes are a little higher, say I'm deliberating whether to ask Nick for primary school recommendations, the mere assumption

---

<sup>52</sup> It also seems relevant that I could have walked down Jubilee Street without the guess: I could have suspended judgment about which street it is, considering how slim my evidence was, and picked a street to try first at random.

that he has children might not suffice for me to proceed. So it looks like assumptions can dispose to action in some circumstances, but in fewer than belief does.

Stakes are one relevant dimension to the likelihood of acting on affirmative attitudes, time pressure is another. If I have one minute to make the train, and it departs either from platform 3 or platform 4, both of which I can reach in just under a minute, I can make a guess or assumption under pressure and act on it. Here, too, it seems accurate to describe the content of my guess or assumption that the train departs from platform 4 as a genuine premise in my practical reasoning.

The defender of right-to-left Belief-Action could try to argue that what is actually being acted on in these cases is a belief in possibility or probability: I am acting on a belief that it is probable that Nick has children, and I am acting on the belief that there is a 50% probability that the train leaves from platform 4. That way, Belief-Action might be compatible with Ex. 4.5 and Ex. 4.6, in which "acting on the guess that p" would just be interpreted as "acting on the belief that p is possible". This kind of reinterpretation of natural language practices should be argued for though. I will say only two things against this potential argument. Firstly, it seems out of line with introspection. Acting on a guess or assumption or conjecture in a low-stakes or high-time-pressure case feels like a natural way to describe our motivational states; reinterpreting them as beliefs in probabilities instead entails a certain amount of introspective error. It is not generally thought to be the case that guesses and assumptions reduce to beliefs about probabilities, they are thought to be mental states of their own kind. On the belief-in-probabilities solution, it would thus look like the same introspective state that in situations where

one is not acting, reveals to one that one has a given guess or assumption, is in cases where one is acting completely misleading, and actually this belief one is not aware of does the work.

Secondly, perhaps more problematically, it might be possible for some creatures that cannot reason about probabilities to have guesses, assumptions, suspicions etc.. Now the Belief-Action defender will probably argue that these beliefs in probabilities should be understood as dispositional beliefs: I am not literally saying to myself in my head "I believe the train is as likely to go from platform 4 as from platform 3", I'm much too rushed for that, but I have the dispositional belief that this is the case and it is that belief that I am treating as true in my behaviour. But even though it is commonly accepted that one can have dispositional beliefs about things one did not consciously entertain (like the probability of the train platform being the right one), it seems implausible that one could have dispositional beliefs about things one *could* not consciously entertain, because one lacks the conceptual or cognitive resources to do so. Therefore, if it is possible for creatures to have weak affirmative attitudes, cases of acting on such weak attitudes are unlikely to be reducible to beliefs about probabilities.

I conclude for now that belief correlates very strongly with disposition to act in both directions, although there are exceptions like certain delusions for the left-to-right direction and low-stakes assumptions and time-pressured or low-stakes decisions in little information for the right-to-left direction. The weaker affirmative states will usually not dispose the subject to treat their content as given, with the above exception. Some affirmative states seem more likely than others to give rise to action in these special circumstances: suspecting, being tempted to believe, and conjecturing seem unlikely candidates, for guessing, assuming, and being inclined to believe it is easier to imagine cases where one genuinely acts on them.

### §4.4.3 Action-inspired normativism

The doxastic normativist can try to put forward a weaker argument from action-disposing, along the following lines. Belief is different from the other affirmative attitudes. This difference is one in disposing towards action: belief disposes very strongly to act on the believed premise, weaker affirmative states do so much less. Action-disposing is intrinsic to belief, and thus belief is subject to an intrinsic epistemic norm. Call this the *revised argument from action-disposing*.

There are a range of problems with this strategy. Firstly, there is the datum that the Strong Belief-Action principle does not hold in psychological reality: some beliefs don't dispose us to action, and some weaker attitudes do. Even if these exceptions to Belief-Action are rare, they make the thesis that action-disposing is intrinsic to belief suspect. Furthermore, once the psychological possibility of exceptions to Belief-Action is on the table, conceptual space opens up for hypothetical creatures for whom these exceptions are more frequent. Even for creatures who act on guesses and assumptions more often than we do, there seems to be something special to the normativity of belief compared to the norms governing the weaker states, and the revised argument cannot account for that. Secondly, there are some propositions that are very unlikely to ever be acted on. Complex mathematics might offer such examples. The intuition that it is more important for beliefs to be epistemically in good shape than for guesses does not seem to water down much for these propositions, whereas that might be expected if action-disposing were the sole reason for the normative difference. Hypothetical creatures who only ever reason about propositions of this sort that are unlikely to come up as premises to act on again seem intuitively to be bound by an epistemic normativity that censures false beliefs. Brains in vats who never act at all also come to mind: their false beliefs are wrong and the

reason for this wrongness does not seem to reduce to negative consequences that *would* arise *were* the brain in the vat act on them, because we can stipulate that the envatted brain in question is part of a population of brains that have been in envatted conditions, utterly incapable of acting for generations, and that furthermore they are envatted in such a way that it is impossible to put them into action-enabling conditions. Under those conditions it would be strange to say that false beliefs get their normative status entirely from what would be the case because of false beliefs for non-envatted subjects.

#### §4.4.4 The function of weak affirmative attitudes

A different source of difficulty for the revised argument is that it leaves it mysterious why the weak affirmative attitudes should be subject to epistemic normativity at all, which they clearly seem to be, albeit a less forceful kind of normativity than what belief is subject to. If belief's normativity stems from its action-disposing character, and the other affirmative attitudes dispose to action only in very exceptional cases, why is it the case that they are wrong when false and irrational when unfounded? The normativist should have something to say about their normativity. Presumably, the weak affirmative attitudes have *some* function in the mind that we can use for this normative explanation. Their role in investigative behaviour comes to mind. Recall Ex.4.4 where a detective is hired as a consequence of one's suspicion. We noted that we can speak of "acting on a suspicion" in such cases, but that the action in consequence is one of launching an investigation, embarking on a course of deliberate evidence gathering. The suspicion in question thus seems to have the epistemic function of stimulating the subject to gather more information. And presumably it is usually better for subjects to launch investigations based on true affirmative attitudes rather than false ones: if I go through life

constantly testing safe water because I suspect it is poisonous and walking down wrong streets to check whether they are the right ones because I guess they are the right ones, that won't be beneficial to me. So the normativist could try a *further revised argument from action-disposing*. Belief is different from other affirmative attitudes. Belief disposes to action and this disposition is intrinsic to belief. Weak affirmative attitudes merely dispose to evidence gathering. Acting on falsehoods is very bad for subjects, acting on weak affirmative attitudes is less bad. Therefore, belief is subject to intrinsic epistemic norms that are more forceful than the norms governing weak affirmative attitudes.

I see three main challenges for the further revised argument, but I will only be able to sketch them. Firstly, the normativist would have to make the notions of acting on *p* and taking investigative steps based on *p* sharp in such a way that it is clear that the latter does not count as an action. Secondly, she will have to get the normative difference exactly right. Hypothetical creatures can again come up for whom the correctness of propositions they could act on is actually not that important (because they rarely act) but for whom for whatever reason not launching investigations based on false premises is very important (police detectives for instance can ruin an investigation by wasting time chasing false leads, we can construe cases where the hypothetical subject finds herself in those kinds of situations very often). Thirdly, the epistemic motivator function of weak affirmative attitudes is more obvious for some kinds than for others. Suspicion seems to match that function very well, as does hypothesis, guessing does so in some circumstances (like making a guess on which street is the right one and then go verify based on that guess), but less so in others, where a guess might close investigation to an

extent<sup>53</sup> (coming to a guess to settle one's uncertainty about a given matter without much intention of actively pursuing the question further). Estimation, conjecture, and assumption seem like states that often close investigation to an extent rather than launch it. Finally, it seems that one can also launch an investigation on suspended belief, complicating the idea that that is the function of our weak affirmative attitudes.

#### §4.6 Against conceptual normativism

The main conclusions of this chapter are as follows. The correctness-based conceptual normativism defended by Shah and Velleman and Engel stands in need of further defence, because the data used to argue for it are misleading (belief is not transparent the way the authors say it is, and neither are the other affirmative attitudes relevantly different in the way at hand) and because the explanation of those data offered by normativism is inconclusive. Secondly, the possibility of disagreement about the norm of belief makes the intrinsic status of that norm implausible. Lawless deviance furthermore makes the intrinsic normativity normatively undesirable. Thirdly, the correctness condition cannot be used to differentiate belief from other affirmative states. On the contrary, the correctness norm makes belief not only normatively but also psycho-ontologically indistinguishable from them. We explored three routes of state-differentiation other than the correctness norm, by which the doxastic normativist might try to protect her view from the problem of psycho-ontological indistinguishability: evidential sensitivity, non-correctness based norms, and belief's action-

---

<sup>53</sup> Belief is often said to close investigation descriptively, whereas knowledge is likely to do so normatively. But one's openness to continuing investigation falls on a continuum, where even though only belief might actually close it in the sense of taking away one's motivation to continue seeking evidence, other attitudes might still go some way towards making one less inquisitive than deliberate agnosticism would.

disposing qualities. Against Shah and Velleman, I concluded that it is not the case that only belief displays “regulation for truth”, and what differences there are between belief and other affirmative states in terms of evidential sensitivity turns out to be difficult to capture in the kind of threshold that would be necessary for a definition of belief. Norms of belief other than correctness still face problems with conceptual disagreement and lawless deviance when we stipulate them to be part of the concept of belief. As long as truth or evidence are the criteria for such alternative intrinsic norms, state-differentiation also remains problematic, and we have seen the dilemma of over- and underprescription arise for such norms. Knowledge norms do substantially better on state-differentiation. Finally, we saw some considerations showing the difficulty of arguing for intrinsic normativity of the concept of belief from belief’s action-disposing qualities in a way that is coherent with the functional and normative role of the weaker affirmative attitudes.

The failures of normativism we have encountered thus far will be important starting points for my own theory of belief in Chapter 7. The normativists proposed certain explananda that normativism is supposed to make sense of. Evidential sensitivity is one of them. The transparency argument is reminiscent of the involuntarism argument that goes back to Williams: belief supposedly displays some form of strong evidential sensitivity, making its formation impossible for the subject to control, and belief’s inherent normativity or telos explains that aspect of belief formation. As we have seen, capturing the actual constraints perceived evidence puts on belief formation is not as easy as the literature sometimes seems to imply it is. Nevertheless, there are some ways in which belief is distinctly sensitive to evidence, there are undeniable ways in which our freedom and control over our affirmative attitudes differs between the different types. And these phenomena, once properly described, should be

explananda for any theory of belief. This is a task I will turn to in Chapter 7. The second explanandum that the normativists identified as important for a theory of belief is that belief is distinct from mental states like guessing and suspecting. Our definition of belief needs to differentiate it from such other mental states. I have sought to demonstrate that this state-differentiation is not easy to achieve in a definition. One reason why a proper differentiation is important particularly in normative contexts, is that normative indistinguishability can lead to over-prescription. Chapter 7 will offer an account of belief that does not struggle with either normative nor psycho-ontological indistinguishability. It will also discuss a way of thinking about the norm of belief on which over-prescription and under-prescription cannot arise.



## Chapter 5: Teleological definitions of belief

### Intentions, aims, instrumental norms

Talk of belief as having an aim has a long philosophical tradition. In the contemporary epistemic norms literature, this notion can be traced back to Elizabeth Anscombe (1957) and Bernard Williams (1970). The former introduced the idea that belief is the kind of thing it is partially in virtue of what it is meant to achieve, the latter used the phrase “aim of belief” to account for our inability to believe at will. As we have seen, accounting for state-differentiation and evidential sensitivity phenomena are still major desiderata for contemporary accounts of belief and its norm. The phrase “aim of belief” continues to be prominent in the literature, but that is not always an indication of teleologism. As Wedgwood, who is not a teleologist but a normativist narrowly construed, writes in an article titled “The Aim of Belief”:

It is often claimed that beliefs aim at the truth. Indeed, this claim has often been thought to express an essential or constitutive feature of belief. But this claim is obviously not literally true. Beliefs are not little archers armed with little bows and arrows: they do not literally “aim” at anything. (p.267)

Many authors, Wedgwood included, think of the aim of belief as a *metaphor* for some feature of it. That metaphor is often interpreted normatively: according to Wedgwood, what it means for belief to aim at the truth, is that belief has an intrinsic correctness norm that makes it so that we ought to believe the truth. Real teleologists in contrast, think that belief genuinely features a purposive element, similar to activities like shooting at a target, warning someone, searching for something, all of which are aim-directed.

## §5.1 What kind of aim?

An interpretation of the slogan that belief aims at the truth that is more in the spirit of teleologism, is interpreting it as being about believers instead of belief. If beliefs aren't little archers that aim at a goal, perhaps they are the arrows. And it is certainly true that we can intend for our beliefs to be true. Of course, the motivations for teleologism about belief extend beyond the interpretation of the slogan that it aims at belief. One of the principal motivations for the view is its treatment of epistemic norms: if believing truly is an aim every believer has, then the normative status of true, false, warranted, and unwarranted beliefs can be straightforwardly explained through instrumental normativity. False beliefs would count as bad because the believer has failed to reach her goal in them. Evidence for beliefs would count as a good thing because it tends to help believers reach their goals. As we've seen in the preceding chapters, the source question about the norm of belief is hotly contested. Teleologism has an attractively simple explanation about where the norm of belief comes from.<sup>54</sup> The simplicity of the instrumental source explanation is one attraction, another is that it is a source of normativity that fits into a naturalist ontology. Here's Steglich-Petersen (2006):

Anyone offering a naturalistic account of belief will tend to prefer the [explanation] involving aims, rather than the [explanation] involving normative judgements. The naturalist avoids commitment to the idea that the metaphysical nature of belief involves normative properties, since it is notoriously difficult to explain these in a naturalistic framework. Aims, on the other hand, seem more palatable for a naturalist. It is usually

---

<sup>54</sup> It can also explain the motivational force of the norm of belief without any of the problematic conflation between norm acceptance and norm adherence that we've seen conceptual normativism struggle with: that considerations of what she needs to do to reach her goals exert motivational force over the subject is much more straightforward than that what she needs to do to fulfil her obligations exerts such motivational force.

assumed that trying to achieve something, having a goal of some sort, and being moved to act by having that goal, can all be explained in a naturalistic spirit. (p. 500)

I share Steglich-Petersen's sympathy for naturalism-friendly accounts of belief norms. Yet the teleological account fails to deliver on that front. In this chapter, I investigate how the teleologist would have to go about defining the aim of belief in order to do justice to the kind of normativity we expect belief to have. I touch on two arguments from the literature against teleologism in the process, and add observations of my own.

First some terminological specifications are in order though. I defined teleologism about belief as the view that "belief is the cognitive attitude a subject has if and only if it is their aim to hold that attitude only if its content is true". That definition is in the spirit of Steglich-Petersen's view, who thinks the primary concept of belief involves a conscious intention on behalf of the believer to believe truly. But there are views that have been called teleological that place less weight on subjects' intentions. Velleman's former theory of belief for instance (Velleman, 2000b) involves a notion of aim-directedness that can refer both to subject intentions as well as to a sub-intentional regulation for truth on the level of belief-producing mechanisms. I stick to the intentional formulation of teleologism partially for brevity's sake in this chapter. My argument against an instrumental reading of belief norms in the second half of the chapter can be adapted to instrumental belief norms generated by sub-intentional teleologism, but I will spell it out using only the intentional version. Furthermore, it is relevant to note that the sense in which sub-intentional truth-regulation is supposedly aim-directed is often put in evolutionary terms: beliefs can be said to aim at the truth because they have been *selected for* aiming at the truth. Since the next chapter is explicitly devoted to problems with evolutionary-teleological accounts, I can focus this one on intentional teleologism.

### §5.1.1 Aim effectiveness

Nishi Shah's (2003) argument from transparency was responsible for converting David Velleman from teleologism (Velleman, 2000a) to doxastic normativism narrowly construed. According to Shah, and the reformed Velleman (Shah & Velleman, 2005), the transparency of deliberative belief is a key explanandum for any theory of belief, and the teleologist is caught in a dilemma when trying to account for it: if she construes the truth-regulation of belief-producing mechanisms strongly enough to explain why belief is transparent in doxastic deliberation, some cognitions like wishful thinking will be excluded from counting as beliefs. The argument from transparency does not necessarily work against a version of teleologism that is construed entirely in intentional terms (Steglich-Petersen, 2006). Furthermore, I argued at length in Chapter 4 that Shah and Velleman's depiction of doxastic deliberation is misleading, and that transparency the way they describe it does not exist.

If transparency is not a real phenomenon, does that let the teleologist off the hook? Not quite. Even if the teleologist were to reject transparency as a psychological necessity, something of a dilemma would remain. Start by considering how much of a difference the intention to believe needs to make to the formation of a belief. If the truth aim is to tell us something interesting about the nature of belief, instead of just telling us some facts about the circumstances in which beliefs occur, the aim will have to have some kind of causal relevance to beliefs as opposed to being merely some intention the believer happens to hold while believing. Consider the activity of birding, which is an aim-defined activity. Some act is only an act of birding if the subject has the intention to see some birds. Whether they actually succeed in seeing any birds on a given birding trip is much less important to the classification of their behaviour than their intention.

But not everything that is accompanied by the intention to see birds is birding. Say I have to decide between Orpheus Lake and Swordfish Lake as destinations for a weekend fishing trip, and my intention of seeing some birds on vacation is part of what makes me decide for Orpheus Lake. While I'm over there fishing and cleaning my boat and collecting firewood, I have in the background the intention to see some birds. But I'm not birding while I'm fishing even if my intention to see birds does make me look up to the sky more than I usually would. To count me as birding the intention to see birds would have to be a much more significant determinant of my behaviour. Nor is it sufficient that an intention to see birds accompanies every single act of fishing of mine to make that act an act of birding, say because I'm someone who constantly has the intention to see birds. And even on a population level happening to always coincide with a certain intention does not suffice to make some act defined by that intention. Imagine for instance that anyone who ever buys binoculars does so with the aim of seeing birds with the binocular. That does not make buying binoculars an activity defined by an aim<sup>55</sup>.

The same is true for belief. Just because belief might always be accompanied by a certain intention of the believer, it does not immediately follow that belief is defined by that intention. If belief is genuinely purposive, the intention needs to make a difference in the formation of the belief, just like one's intention to see birds needs to make a substantial difference to one's behaviour in order to count as birding. Now take a case like the following:

Ex. 5.1: I'm thinking about whether my brother lied to me about the size of my inheritance. In trying to make up my mind, my hatred of my brother which incentivises

---

<sup>55</sup> One might think buying binoculars is defined by the aim to obtain binoculars, but that is not true. One *only* counts as buying binoculars if one ends up with binoculars, and one can count as having bought binoculars even though one did not have that intention (one accidentally bought them). Buying binoculars is result-defined. Contrast this with looking for binoculars, which is aim-defined.

me to think badly of him is one motivating force. The fact that I already spent my allocated inheritance and would really like it if there turned out to be more to inherit also plays a role: I am driven to believe something pleasant, and something that will be instrumental in my attempts to get more money. Finally, the evidence that my parents' estate did seem a lot larger than what my inheritance would suggest is relevant to my intention to believe something true.

In Ex. 5.1, I have an aim for my belief that it be true. I intend to hold the belief only if it is true. But this intention does not play a major role in the formation of the belief: its motivating force is only one of several that are relevant to the belief in question. So the intention to believe truly seems like it might be *external* to the belief the way the intention to see birds is external to my act of fishing above. If that is the case, the teleologist faces a dilemma. On the one hand, if she defines the truth aim in such a way that it makes sufficient difference to belief formation so that the aim not merely *accompanies* belief but *defines* it, cognitions like the one in Ex. 5.1 will not count as beliefs. On the other hand, if she wants to prevent that consequence of state exclusion and count the cognition as a belief, arguing that beliefs are accompanied by an intention for them to be true does not actually suffice to show that belief is a purposive state.

### §5.1.2 Norm tempering

There is a reason why the teleologist should opt for the first horn of the dilemma, and hence for a theory that includes state exclusions. The aim of belief is supposed to generate epistemic normativity: beliefs would get their specific normative statuses from their contribution to the subject's aim. But for that picture to work, the teleologist needs to define the intention accompanying beliefs strongly. Here's why. Teleologism is meant to explain why false beliefs

are wrong. But our beliefs can contribute to multiple aims: they can succeed or fail in helping us recover from alcoholism, they can be helpful or unhelpful in mustering up confidence, they can bolster or compromise our loyalty to our superior. But it is *only* their contribution to the truth aim that determines their *wrongness*. The intrinsic aim hypothesis can handle this. To see how, consider the example of searching, which is an aim-directed activity. Say I go to the forest to search for mushrooms. Like birding, whether I am mushroom-searching is determined by whether I *intend* to find mushrooms, not by whether I succeed. Say I'm mushroom-searching but also have the aim to see some birds. For instance, when I have to decide between two areas to search that seem equally likely to contain mushrooms, I will choose the one that I think more likely to also contain birds. Now I'm back home from mushroom-searching and didn't find any mushrooms, although I did see two woodpeckers. My mushroom-searching was *unsuccessful*. This normative status is entirely dependent on whether I found mushrooms or not. That the mushroom-searching did further some of my other aims I had whilst mushroom-searching is irrelevant.

So far so good. But now the analogy between searching and believing breaks down for two reasons. Firstly, because it's perfectly permissible to go on unsuccessful mushroom searches. Secondly, because the overall normative status of the failed mushroom search is *tempered* by the fact that it contributed to other goals of mine: my failure on the one account is alleviated by my success on the other. The badness of wrong beliefs however seems not remedied by the contribution the wrong belief made to other aims of the agent (like making her happy, or fuelling her anger at someone). Indeed, it seems to be made worse by it. If the normativity of belief came entirely from agent intentions, we would expect a belief that fails to meet the agent's

truth aim but fulfils many of her other desires to be a good belief, but the opposite seems to be the case.

What this illustrates, is that the teleologist can only hope to explain the normativity of belief in instrumental terms if she defines the intentions to believe truly that are meant to accompany beliefs rather intensely: it's important for the subject to not have too many other goals and motivations at play when forming beliefs, and that the truth-aim plays a significant causal role in their formation. But unpacked like that, the intrinsic aim hypothesis sounds less plausible. It would for instance entail that one can go from believing to not-believing simply by having one's other intentions for that belief increase in intensity a little. And of course, that will mean that the cognition in Ex. 5.1 for instance will probably fall below the threshold of aim proportion.

### §5.1.3 Telos by extension

The discussion of aim intensity above, and the related question of the classification of cognitions that contribute to non-alethic aims (like wishful thinking), has as a premise that every single belief of a given believer needs to be formed with the right intentions. But maybe the truth aim does not need to accompany *each* belief. Steglich-Petersen thinks it does not:

So the primary concept of believing *p* is that of accepting *p* with the aim of doing so only if *p* is true... Furthermore, cognitive states and processes that are not connected with any literal aim or intention of a believer can nevertheless count as 'beliefs' in virtue of sharing precisely the descriptive characteristic of being to some degree conducive to the hypothetical aim of someone intending to form a belief in the primary strong sense. (2006, p. 515)

Steglich-Petersen's suggestion is that some mental states that are not accompanied by a truth-aim can get the status of belief by extension from those mental states that are accompanied by it, in virtue of sharing other, non-teleological properties with them. But this explanation does

not fit at all with how we think about standard examples of aim-directed activities. Steglich-Petersen uses the example of concealing as an aim-directed activity to draw analogies with belief. This is strange because concealing seems like a straightforward example of a result-defined activity rather than an intention-defined one: one can fail to conceal something even though one meant to conceal it and used a method that counts as broadly “conducive” to that goal, and one can also conceal something without intending to. I will discuss two examples of aim-directed activities that I think are much more obviously purposive than concealing is, in order to show that what Steglich-Petersen predicts for belief is not the case there.

Take first the example of shooting at sparrows. This is an aim-directed activity because the intention to shoot a sparrow is necessary for a shot to count as shooting at sparrows: if I mean to shoot at pigeons but there are some sparrows in the sky as well, then I am not shooting at sparrows. If I mean to shoot a pigeon but aim poorly and hit a sparrow instead, I shot a sparrow but not through the activity of shooting at sparrows. The intention is probably not sufficient to count as shooting at sparrows (for instance if I try to shoot at sparrows by shooting at pigeons because I think that pigeons are sparrows in disguise, that might not qualify as shooting at sparrows<sup>56</sup>). Nevertheless, the intention seems to play a major role in determining whether some shooting behaviour counts as shooting at sparrows.

Now recall Steglich-Petersen’s suggestion: mental states that are like aim-directed beliefs in their descriptive properties and are conducive to the aim the subject would have had had she aimed at the truth can inherit the status of belief by extension. Could something like that be true of shooting at sparrows? Can a shot that I did not at all intend to hit a sparrow qualify as

---

<sup>56</sup> In the *de re* sense of “shooting at sparrows” as opposed to some possible *de dicto* sense.

shooting at sparrows in virtue of being conducive to sparrow-hitting? Say I intend to shoot sparrow hawks in order to protect sparrows. But my evil demon tries to thwart me by making all my attempts to protect sparrows harmful to sparrows, so whenever I shoot at a sparrow hawk my bullet actually hits a sparrow. In that scenario, what happens when I try to shoot a sparrow hawk is very reliably conducive to the aim (that I don't have) of shooting sparrows. Does the conduciveness to the hypothetical aim make what I do shooting at sparrows? It seems not.

Take another example of an aim-directed activity: practicing  $\Phi$ , where  $\Phi$  is something like badminton, the Partita in D minor, rope walking. It seems that something only counts as practicing  $\Phi$  if it is intended to improve or maintain a skill relevant to  $\Phi$ -ing. If my friends have never heard of rope walking and I demonstrate a rope walk for them with no intention regarding the development of my own skill but just to convey the basic concept, that instance of rope walking does not count as practicing rope walking. We see this necessity of the right intention in order for a token action to count as practicing  $\Phi$ , such that something being done with the right intention cannot be inherited through descriptive analogy, even when we stipulate the conduciveness to a hypothetical aim explicitly. If I play my grandmother a piece on the violin as poorly as possible in order to make her give up on wanting me to learn to play, but it turns out that by playing poorly on purpose I actually improve my violin playing skills, the conduciveness of this exercise to the hypothetical aim of improving does not make it an instance of practicing. Reflection about aim-defined activities does not support Steglich-Petersen's suggestion that beliefs, even though the primary concept of them is defined by an intention, can nevertheless fail to be accompanied by this intention since they can count as beliefs by extension from the intentioned ones.

It also bears mentioning that even if the picture of non-truth-aimed beliefs counting as beliefs works, that does not yet mean that these by-association beliefs will also have instrumental norms governing them that say that they ought to be true. That would be a very surprising result. If belief's general normativity comes from the subject's aims, so that mysterious sources of epistemic normativity can be avoided, but some beliefs do not have this direct aim-generated normativity, why should there suddenly be a different, non-direct-aim-generated norm for those beliefs just because they fall under the same category in virtue of their descriptive similarity to purposive beliefs? Belief is a normative phenomenon on teleologism *only because* of the intentions that accompany it. It would be strange if for states that lack this aim the subject's intentions or lack thereof could be overridden by a normativity stemming solely from their classification as beliefs.

Telos-by-extension is an unpromising avenue. If belief is indeed intrinsically intentional or aim-directed, *all* token beliefs will have to meet the condition of being accompanied by the right aim, where the right aim means that among the subject's intentions for a given cognition, her intention for that cognition to be true needs to be significantly stronger than competing intentions she may have. And on that construal of the truth aim, we see cases of lawless deviance arise: subjects whose truth aim competes with too many other aims, so that they cannot be said to aim at the truth *sufficiently*, will be disqualified from belief, and hence from the norm of belief. In so far as they have the aim to believe truly *to an extent*, their false beliefs will still count as wrong to *that* extent, but their overall normative status may well be a positive one if the failure to be true is outweighed by successes with respect to other aims. What all of this illustrates is that one of the major flaws of teleologism lies ironically in one of its main advertising points: that we can supposedly explain doxastic normativity easily, and non-

mysteriously, through instrumental norms. But the norms that govern our beliefs are structurally dissimilar from instrumental norms. I will try to show that this dissimilarity is substantial.

### §5.2 Instrumental belief norms

Instrumental normativity is in a sense private. One is subject to much less public censure for failing to do something one had to do only by one's own lights. When I say: "Frank wanted to buy organic aubergines but he didn't" in a context where only Frank wanted organic aubergines, the amount of criticism I can bestow on Frank is limited. I can criticise him to the extent that his failure has negative effects on me (maybe Frank is complaining all night about the lack of organic aubergines, or maybe I don't get to eat any aubergines now where I wouldn't have cared myself whether they were organic). I can also criticise him if the failure manifests a character flaw: maybe this is indicative of Frank often being careless, or maybe I simply think people who don't reach their own goals are losers. I might also empathise with Frank in a way that makes me feel his disappointment as my own. But outside of these cases, telling someone "you didn't obtain a goal you had set for yourself" is much less censuring than telling them they violated a non-instrumental norm.

Now the reader may think that the above analogy is unfair to the teleologist, because the goal of wanting organic aubergine is so idiosyncratic compared to the goal of having true beliefs. But how widely shared the goal is has no bearing on the force of instrumental normativity. Take the goal of being happy. Almost all humans share this goal, although with wildly differing conceptions of what happiness consists in or how to go about reaching it. Therefore, in most ordinary contexts, when we talk of someone as unhappy, it is understood that this means that person fails to obtain something that she wants. Happiness is thus much like the truth aim that

the teleologist argues is understood to be had by believers. But calling someone unhappy is not ascribing a mistake to them, whereas calling someone's belief false is. The difference is especially clear when the goals of the ascriber and the evaluated subjects differ. Say I want Georgia to be unhappy (I hate Georgia), and you do too. When I say to you "Georgia got broken up with again, she's really unhappy", calling Georgia unhappy does not involve ascribing fault to her, even though it is clear that Georgia is failing to reach an important goal of hers (and one that is near-universally shared). Now imagine that I also want Georgia to have false beliefs. Maybe I enjoy watching Georgia embarrass herself through her misinformed assertions. If I say to you "Georgia studied for the politics test with a book from 1992, and it made her believe Helmut Kohl is the chancellor of Germany", I am clearly calling Georgia mistaken, even though the mistake is one I want her to make. Being unhappy however is not a mistake, even though it is a failure to reach a goal one has almost by definition. It seems that the faults we ascribe to others when we call their beliefs false are similar to the faults we ascribe to others when we call their actions reprehensible or their attire ugly: regardless of whether they want to be good or well-dressed, we are calling them in the wrong with evaluations like reprehensible or ugly. Calling someone unhappy however communicates criticism only by the lights of the unhappy person's own desires, and by proxy my empathy with their desire. This is why, when my desires and those of the evaluated person come apart, my ascribing unhappiness to them does not communicate fault. The normative dimension of ascribing false belief goes much further than that of ascribing failure to have reached a goal.

### §5.2.1 Optional normativity

The kind of criticism they open one up to is one aspect of instrumental norms; the most crucial one is that they can be opted out of: an instrumental norm to do  $\Phi$  in order to  $\Psi$  only applies to one in so far as  $\Psi$  is a goal one has. Of course, teleologism about belief does not just make belief norms conditional on some goal believers may happen or not happen to have; they are conditional on a goal all believers by definition have. But from it being the case that all believers have the truth aim it does not follow that all epistemic subjects have the truth aim. An epistemic subject is something that is subject to epistemic norms. If some epistemic subject fails to have the truth aim, on teleologism that means she is not a believer. Hence, no belief norms censuring her false propositional attitudes apply to her. And this is the problem of lawless deviance once more, indeed the worst manifestation of it we have seen thus far: on an instrumental belief norm, if the subject does not intend for her belief to be true, a false belief will not count as wrong at all. Let's look at some cases where affirmative attitudes are not intended to be true by their subjects.

Ex. 5.2: Marie is waiting for her blood test results in the mail which she thinks will be negative, but is worried will be positive, and now she gets an envelope in the mail that looks like it might contain the results. Marie says to herself "It's probably just a bank statement, I can open it later" and puts the envelope on her desk instead of opening it.

Marie has an affirmative attitude that the envelope is not important and can be opened at some later point. It's plausible to describe Marie as not aiming at the truth with this attitude: nothing in her present evidence makes it more likely that the envelope contains something that can be dealt with later rather than that it contains the test result she needs to know urgently, and

opening the envelope to acquire such evidence would take minimal effort. Her only reason for thinking it's a bank statement is her anxiety about the test results: deciding that she does not need to open this envelope now enables her to postpone the moment where she has to confront what she is scared of. So Marie's intention behind her affirmative attitude, the goal she is aiming at with it, can be more adequately described as protecting her mind from the risk of unpleasant information than as representing the word the way it is.

### §5.2.2 Derived instrumental norms against lawless deviance

Ex. 5.2 is a case where an epistemic subject fails to aim at the truth with a single belief. On Steglich-Petersen's behalf, one might try to prevent optional epistemic norms via the proviso of aim-directedness by extension. Recall that on his account, if an attitude has some minimal conduciveness<sup>57</sup> to the hypothetical aim of believing what is true, that descriptive similarity to the truth-aimed beliefs suffices for its status of belief-by-association. I already discussed multiple problems for the telos-by-extension hypothesis in §5.1.3, but if this hypothesis could prevent the disastrous result of belief norms being something we can opt out of, the teleologist might want to stick to it anyway. Let's imagine then that telos-by-extension can be made more plausible, and investigate whether it can prevent lawless deviance.

---

<sup>57</sup> In so far as we even accept that Marie's cognition meets the non-teleological criterion of conduciveness. Presumably, there will be some cases of wishful thinking that meet the minimal-conduciveness condition, but very plausibly there will be others that do not. For these, the teleologist may have to accept the consequence of state exclusion. That is a case of lawless deviance in itself of course: the worst cases of wishful thinking get no censure by epistemic norms, but the less bad ones do (for they make the cut to count as beliefs). Whether the teleologist can ultimately account for wishful thinking via the belief-by-association route is a matter I will leave open. If truth-conduciveness qua descriptive feature of propositional attitudes suffices, perhaps other non-teleological properties of belief, like confidence or disposing to action, will too, and we can count wishful thinking as belief that way.

Say we grant that Marie's affirmative attitude about the envelope can count as aim-directed by extension: it is descriptively similar to her normal beliefs. Those normal beliefs fall under the primary concept of belief because they are aim-directed. But now the question opens up what warrants this primary concept of belief. Is it the kind of belief-like state the agent has more of, so that candidate-beliefs that do not fall under the primary concept need to be exceptions in her psychology? In that case, trouble comes from the conceptual possibility of beings who have the wrong proportion of truth-aimed to non-truth-aimed beliefs: in such beings belief would be impossible: their truth-aiming states will not count as beliefs in the "strong primary sense" because there are too few of them, and the non-truth-aimed states cannot be counted as beliefs by extension as a consequence. Of course, it would make more sense to construe our reasons for conceiving of belief under the strong aim-directed concept on the level of a population instead of within one individual's psychology. But again, that believing beings on population-average have more states that are truth-aim-accompanied than not is *contingent*, and the concept of belief should not be depending on it.

A different way of thinking about warrant for the strong primary concept would be through stereotypes: what our stereotype concept of a dog looks like (medium-sized, short-to-medium hair, probably brown) seems to not depend on whether in fact there are more chihuahuas and chow-chows than there are dogs resembling our stereotype. But if teleologism is just the defence of our concept of belief as aim-directed the way our concept of dog is brown, it doesn't seem like an interesting philosophical view. As philosophers, we use the concept of belief hoping to generate true conclusions about which mental states we should classify how – in order to do that kind of work, that concept needs to be in better standing than a stereotype.

Telos-by-extension, even if it could be made plausible against the considerations presented in §5.1.3, cannot save teleologism from lawless deviance. Trying to censure cognitions that the subject does not intend to hold only if they are true through the descriptive similarities between those cognitions and some primary concept of belief is dependent on the force of that primary concept. If the subject in question simply does not share Steglich-Petersen's concept of belief as a state that is aim-directed in its primary sense, it is unclear where the instrumental normativity governing her deviant mental states is meant to come from. If we want to argue that belief norms nevertheless apply to such a subject because most of her beliefs happen to be accompanied by the intention that they be true, we operate with a concept of belief that is contingent on the proportion of truth-aiming to non-truth-aiming mental states a given subject has or that exist within the population at large. This opens up the possibility for creatures who have no beliefs whatsoever on teleologism due to their problematic aim distribution. Arguing that non-truth-aiming cognitions can count as intentional-by-extension requires a plausible account of the concept of belief that this extension is supposed to originate from, and why the concept should apply to subjects without truth aims. The avenues to provide that kind of account we briefly investigated did not look promising.

### §5.3 Against intentional teleologism

We have noted three groups of problems for intentional teleologism. The first stems from the question how to interpret the intrinsic aim hypothesis: if belief is by definition accompanied by an intentional truth aim, what position should that intention occupy among the subject's overall aims? In order to explain our normative intuitions about scenarios where false beliefs fulfil non-epistemic aims (they are beliefs the subject should not have and their wrongness is not

remedied by the goals they do contribute to), we need to define the truth aim as occupying the lion's share of the subject's overall intentions regarding a given belief. That leads to a first set of state exclusions, and lawless deviance for those states.

The second group of problems stems from the nature of instrumental normativity: the kind of censure we expect failure to meet instrumental norms to open agents up to is much weaker than the censure we are used to performing for failure to meet epistemic norms. Additionally, instrumental norms can typically be opted out of, which seems highly unintuitive for the norm of belief. The implausibility of opting out of epistemic normativity goes right to the heart of the lawless deviance problem, which makes for the third group of problems that we've seen plaguing teleologism on multiple fronts. It is an unfortunate consequence of instrumental belief normativity that cognitions that fail to be accompanied by a truth aim fall outside the scope of the belief norm. This form of lawless deviance seems to be the worst one we have encountered until now. It opens up the possibility of a deliberate opting out of epistemic normativity through exactly that trait that our folk-morality condemns most confidently: bad intentions.

## Chapter 6: Teleo-functional definitions of belief

### Etio-teleology, functional monism, function conflicts, deviant histories

Thus far, I have looked at three types of normativist (broadly construed) accounts of belief. The first is state normativism, which defines belief as a mental state with normative properties. The second is conceptual normativism, which argues that our way of thinking about belief necessarily involves normative notions. The third is teleologism, which defines belief as the mental state that is produced by the aim of believing the truth, as a result of a conscious intention to believe the truth. I have rejected all of these. Recall that what makes all these accounts normativist (broadly construed) on my definition, is that normative properties of belief follow from the definition of belief. For the teleologist this comes in the shape of instrumental normativity resulting from the believer's aims. For normativism narrowly construed this is in the form of a categorical norm of belief, like the correctness condition or more explicitly prescriptive/restrictive deontic norms. The only proposal of an evaluative belief normativism that I will study in detail in this thesis is teleo-functionalism.

#### §6.1 Functions and norms

Functions play an explanatory role in many areas of contemporary philosophical thought. They are a key concept in the philosophy of biology, but also, through the teleosemantic project, of relevance to philosophy of language and philosophy of mind. They are increasingly appealed to in epistemology as well. There is Plantinga's (1993) account of warrant and Bergmann's (2006) functionalist account of justification, Copp's functionalist account of epistemic

normativity (Copp, 2009), and Kelp's functionalist account of assertion (Kelp, forthcoming). Belief has been defined through functions by the teleosemanticists already (Millikan, 1984, 1986), and this is now also happening in the epistemic norms literature (Graham, 2014b; Nolfi, 2015; Papineau, 2013).

Attempts to define belief through its presumed function are normative in a far less thick way than the other theories we have studied. They involve neither deontic norms nor intentional aims. Of course, teleo-functionalism has a certain teleological flavour, as we can talk of functions being "fulfilled" or processes being "aimed" at, but this is metaphorical language compared to the intentional teleologism we encountered in Chapter 5, where goals are assigned to believers. Yet even teleo-functionalism entails belief norms: from an item having a certain function, conditions follow for its functioning correctly, or being successful. These evaluations do not automatically translate into obligations or permissions, so one would need a bridge premise that says something along the lines of "you ought to have such beliefs that your belief-producing mechanism fulfils its proper function" in order to derive a thick norm of belief from belief's proper function. Nevertheless, even the limited normativity that is intrinsic to the nature of belief on teleo-functionalism will be rejected in this thesis.

Teleo-functionalism construes belief by reference to a function which is defined in terms of what is normally<sup>58</sup> the case. Papineau (2013) writes:

Some readers might be wondering whether this kind of biological account is really an alternative to the view that the nature of belief depends on norms. After all, biological functions can be equated with activities that are 'designed' or 'supposed' to occur, and to this extent can be viewed as specifying 'norms' of functioning. However, while I have

---

<sup>58</sup> The notion of "normal" on etio-functionalism is a special one, that does not reduce to what is usually the case. I explain the details below.

no great objection to this kind of terminology, it is not the kind of norm that is at issue in this paper. We are here concerned with norms that have prescriptive force—that indicate what *ought* to be done. Purely biological ‘norms’ have no such implications. (p. 73)

For Papineau, functions are not real norms at all. This is different on my view. I count the evaluations we can apply to biological entities reflecting their accordance with their function as normative, albeit less thickly normative than prescriptions and other evaluations like “good”. Reflecting about artefact functions gives us cause to think of functional evaluations as normative judgments. For instance, if I sort a pile of cell phones into functioning and non-functioning ones, there seems to be a clear sense in which that classification is more normative in nature than one by which I sort a pile of rocks into igneous and non-igneous ones.<sup>59</sup> The normativity of “functioning” is what makes teleo-functionalism about belief a species of doxastic normativism (broadly considered.) Yet even if one disagrees with the notion of a function-evaluation as normative, there are reasons to treat functionalist accounts of belief alongside the three more obvious candidates above.

---

<sup>59</sup> When applied to biological entities, judgments about the normative status of selectional histories seem to vary to some extent depending on the biological entity in question. We seem happy to classify our judgments about hearts as functioning or non-functioning as normative ones, and even to say that there is a sense in which a heart “ought” to pump blood. But when we look at evolutionary theories according to which human fists have evolved to fight other humans, or according to which hidden oestrus in human females evolved to aid female promiscuity, the classifications of well-hidden oestruses or fists well-suited for punching might strike one as less normatively endorsing. As always with normative judgments, there is a risk that our moral judgments swamp our other ones. Cuckolding and fist fighting have negative moral connotations, surviving has a positive moral connotation, and hence endorsing heart functioning in a positive normal sense gives rise to less inter-normative conflict than endorsing fist functioning. We might have some subconscious aversion to normative judgments that will give rise to second-order normative conflict or dilemma. For that reason, I will not pay much heed to the potential intuition that nothing normative follows from the fist’s selection history vs from the heart’s selection history. I will assume that teleo-functionalism commits one to certain thinly normative evaluative judgments about all biological entities that have functions.

Firstly, functionalists are interested in success conditions for belief, much like normativists about the nature or concept of belief. For the former, these are the conditions that explain belief's existence throughout evolutionary time. For the latter these are the conditions under which a belief gets a positive evaluation of some sort. Findings about markers of evolutionary "success" - even if one uses this notion non-normatively - influence our evaluative judgments.<sup>60</sup> So getting a sense of what biological "success" conditions might be, is of interest to many people when thinking about belief norms. Secondly, the kinds of success conditions teleo-functionalists and normativists (narrowly construed) have in mind are very similar (e.g. truth, evidence). Thirdly, teleo-functionalism uses these success conditions for a definition of belief, just like normativism does. As we will see, some of the problems with teleo-functionalist definitions of belief are common to the general normativist strategy of defining belief. Fourthly, appeal to proper function has been explicitly used in the literature to argue for specific belief norms. Kate Nolfi (2015), for instance, appeals to the notion of belief's proper function in a paper titled "How To Be A Normativist About Belief".

## §6.2 Preliminaries of etiological teleo-functionalism

It is important to clarify two distinct ways of using the word "function". One is purely descriptive and has to do with describing a certain process causally from input to output. This is the notion of "function" at hand were I to ask you "How does this bicycle function?" and you would explain to me the causal chain that leads from the input of pedalling to the output of turning wheels. The other, more normative notion of "function" is at work were I to ask, "What is the function

---

<sup>60</sup> E.g. the marked trend in folk-thinking about gender differences to use so-called findings about evolutionary "reasons" for gendered traits (male promiscuity, female maternal instincts) as vindications of the continued promotion of those traits.

of the chain in the bicycle?" and you would explain to me what the chain is supposed to do and how it contributes to the goal of getting the bicycle to move. A crucial difference between the purely causal *functioning* (verb), and the more normative *function* and *functioning* (adjective) is revealed when we think about broken bicycles: you can explain to me how a broken bicycle functions (in the causal sense) without classifying it as broken and without mentioning how it *should* function. The function (in the normative sense) of a broken bicycle chain however is still to get the bicycle to move, so to explain its function one would have to make mention of how it would behave in a bicycle that is not broken. The two modes of explanation interact: understanding what the (normative) function of a given artefact or entity is can be instrumental to uncovering how it functions on the descriptive level.

#### §6.2.1 Functions and functioning

A functioning and non-functioning bicycle function differently. Yet they are both bicycles, and they both have the same function in the normative sense. For that reason, when we talk about bicycles on the type level instead of on the token level, we need to make reference to its function in the normative sense, since we want our type 'bicycle' to extend to non-functioning token bicycles. This explains the interest in a normative kind of function of the mind: if we're interested in defining types of mental states, simply describing the causal chain from their inputs to their outputs won't suffice, since some token mental states might function differently from others, and then we will need an explanation of why they are nevertheless of the same type. The failure to account for malfunctioning through dispositional functionalism was a major inspiration for teleosemantics. But a similar challenge inspires normativism about belief. Unless we find descriptive elements that all beliefs (including malfunctioning ones) have in common, we need

some explanation as to why the descriptively different tokens are all of the same type. The function of belief is supposed to provide that kind of explanation.

The kinds of functions I will examine in this chapter are etiological ones, bracketing other accounts like systemic-capacity functions, as well as accounts of artefact functions. This choice is made mostly for reasons of space, and partially because etiological/evolutionary accounts of a proper function of belief seem to enjoy popularity in the contemporary epistemic norms literature.<sup>61</sup> The etiological notion of function goes back to Larry Wright (1973), who defined functions as consequences that explain why a certain trait exists. The etiological project is backward looking through-and-through. Papineau, for instance, writes: "As I understand claims about biological functions, they are equivalent to claims about selectional histories" (p.73). From now on, when I write "functionalism", it is short for "etiological teleo-functionalism". I will argue against the project of defining belief by reference to an etiological proper function.

A belief, on functionalism, is the kind of mental state that is produced by something that in the believer's ancestors tended, under normal conditions, to produce true beliefs, and where the current existence of the belief-producing thing is explained by the positive contribution it made to the ancestor's fitness. The awkwardness of talking about "the belief-producing thing" is deliberate. Most authors don't go into detail about the nature of the belief-producing entity they have in mind for which they are defining etiological functions. When Graham (2014), for instance, writes about "the belief-producing process", is this process physiological, or

---

<sup>61</sup> Artefact functions seem *prima facie* difficult to apply to beliefs save by analogy, so the hypothetical project of warranting a robust belief norm by reference to artefact functions is not one we need to engage with.

behavioural, or something different? And if the belief-producing process is meant to be physiological, on what level of biological organisation is it found?

### §6.2.2 Biological items

Graham, along with i.a. Millikan (1984) and Nolfi (2015) uses the example of the heart and its function of pumping blood to elucidate his account. Whatever thing one imagines being responsible for the production of belief, it is unlikely to be an organ. So, hearts and belief-producing entities are presumably very different things, biologically speaking. Given that these functionalists are looking to biology for the notion of function that they are hoping to apply to belief, lack of clarity about which units they have in mind for those functions is potentially problematic, because functions are studied on many different levels in biology. A single amino acid has a function in the primary protein structure (peptide links to other amino acids to form a sequence) but can also have functions in the secondary (i.a. hydrogen bonds to other amino acids to form a alpha helix) and tertiary protein structure (i.a. sulphur bridges to other amino acids to hold a folded shape), as well as functions on the level of the whole cell, like forming the ATP-binding pocket of the sodium-potassium pump. Across all these different levels of organisation, there is no well-defined single notion of function.<sup>62</sup> How should we then think about the biological units that are the carriers of etiological functions, so that they include both hearts as well as belief-producing things? A frequently recurring word among functionalists for the kind of thing that has a function is "trait", as in the following quotes by Millikan (1993), McLaughlin (2000) and Graham (2014b) respectively:

---

<sup>62</sup> This may simply be because very little biological work is being done that is concerned with more than one or two levels of organisation.

Only if an item or trait has been selected for reproduction, as over against other traits, because it sometimes has a certain effect does that effect count as a function. (pp. 35–36).

[T]he distinction between effects of a trait that are functions and effects of a trait that are simply coincidental is fundamental to the notion of function. (p.68)

Maintenance selection, on the other hand, maintains the frequency of traits in a population. Most mutations, for example, are harmful. [...] Because harmful they are selected out, in favor of the normal variant of the trait. The non-malformed trait then continues to exist and predominate in the population because of its relative superiority. (p.12)

Unfortunately, in biological literature, even the word “trait” is used rather ambiguously. On one usage of “trait” (also “character”, also “phenotypic trait”), traits are found on the species level, and can be expressed in individual organisms through differing “aspects” (Martin & Hine, 2008). On that coarse-grained notion, having hair is a trait, but having blonde hair is an aspect. On the other usage, traits are found on the level of the organism. On that fine-grained notion, having blonde hair is a trait itself. Other terms sometimes used for traits on the organism level are “Mendelian trait” and “functional trait” (Martin & Hine, 2008). For the functionalist project, we should use the fine-grained notion. This is because aspects of one coarse-grained trait in a single species can clearly have different functions. Feather colour in female pheasants has a camouflage function, feather colour in male pheasants has a function of visually appealing to females. In humans dark skin colour may have the function of protecting against UV radiation, whereas light skin colour may have the function of allowing more vitamin D production.

With these considerations in mind, I will henceforth use the word “trait” to refer to the unit for which we study etiological functions, where a trait is any characteristic of a phenotype, regardless of whether this trait is shared by all healthy individuals of the species (having a heart)

or whether it is polymorphic (having dark skin). In the definition “any characteristic of a phenotype”<sup>63</sup>, both “characteristic” and “phenotype” are to be understood maximally generally. Characteristics are those observable by any means, from protein immunoblotting for macromolecule detection to counting eggs in bird nests over a period of years to detect average brood size. Phenotype is thus also understood as encompassing all levels of biological organisation, from an amino acid to a whole organism, as well as extensions of the organism, such as its behaviour and the products of its behaviour (like nests). This broad definition should allow us to capture the belief-related phenomena for which we want to investigate etiological function of outstanding empirical results about the nature of belief.

The extended-phenotype approach also circumvents nature versus nurture discussions about belief. Although it seems unlikely that what disposes one towards more true beliefs than false ones should turn out to be a matter of entirely learned qualities<sup>64</sup>, since behaviour counts as a part of the phenotype, we can apply the etiological analysis to traits that are not genetically determined. To warrant investigation into belief-producing structures’ etiological function, the functionalist needs another assumption: that the traits in question are more-or-less direct products of non-random selection of some sort. Non-random selection can include natural, artificial, and sexual selection, but will crucially exclude genetic drift. The kinds of events that fall

---

<sup>63</sup> One might worry that this definition of trait leaves open the possibility of treating conjunctions of disparate traits as a trait such as “being female and having black hair and blood group O and being colour blind”. I agree that it would be strange to call random conjuncts of properties “trait”, but this needs not worry us. No one is arguing after all that *all traits* need to have a function, the conjunctive traits will not pose a counterexample.

<sup>64</sup> Because some of the skills we know of that affect one’s adeptness at forming true beliefs the genetic basis of which we have investigated have turned out to be (at least in part) genetically determined: colour vision, auditory capacities like absolute pitch, memory capacities like eidetiks, or the belief manipulation capacities captured under the concept of intelligence.

under genetic drift, like bottlenecks, can in a sense be said to select certain organisms for survival, and therefore to cause persistence of certain traits. But the random nature of genetic drift means that this “selection” lacks the kind of explanatory connection between trait and persistence that is required under the etiological definition.<sup>65</sup> Being a more-or-less direct product of selection excludes evolutionary spandrels from having etiological functions for the same reason: traits that are spin-off effects of the selection of other traits lack explanatory connection between the trait and its persistence.<sup>66</sup>

In sum, the teleo-functional account of belief assumes that belief-producing traits have been non-randomly, more-or-less-directly, selected for. Given our limited knowledge about the physiological nature of belief, our knowledge of its evolutionary history is even more limited. However, the ease with which direct adaptational benefits of certain cognitive behaviours come to mind, and the fact that in biology, adaptational explanations usually have the benefit of the doubt over spin-off ones makes assuming more-or-less direct, non-random selection *prima facie* fair. We now have the vocabulary in place to study the function of belief-producing entities. The possession of those entities are traits of individual phenotypes which we assume to be the product of non-random more-or-less direct selection.

---

<sup>65</sup> In a bottleneck event, like an earthquake, the population of a given species is reduced drastically, sometimes nearing the minimal viable population size. A small number of organisms survives. If some bird species on an island exists in dark and light morphs in roughly equal numbers, and after a bottleneck hurricane the only 10 birds left all happen to be dark morphs, the future population on that island will be predominantly dark, but this has nothing to do with the property of being dark.

<sup>66</sup> Concrete selectional histories very often contain both selection as well as drift. The emphasis on being selected *for* in the definition of etiological function might make for difficulty regarding traits with so-called “piggyback selection”. This is a potential shortcoming of teleo-functionalism I will not have space to explore.

### §6.3 Belief as function-constituted

Here are some examples of accounts on which belief's proper function, whatever it may be, is defining of belief. Millikan (1984) for instance, sees etiological functions as the answer to the problem of formulating biological categories that are abnormality-inclusive:

For example, just as hearts and kidneys are sometime diseased or malformed, so sentences and beliefs are sometimes false ... Such sentences, beliefs, words, and concepts are not able to perform their proper functions. Consider, for example, beliefs. The traditional functionalist has assumed that whatever interesting functions make beliefs to be beliefs, these must be found among the actual dispositions or powers of beliefs or of people who have beliefs. True beliefs are beliefs; false beliefs are also beliefs. So false beliefs must have these powers too. But this, I will argue, is like looking for the functions that make kidneys to be kidneys by trying to see what is in common to the powers of all kidneys – the normal ones, the diseased ones, and the malformed ones. ... But if, instead, we push the analogy with biological categories, only true beliefs are capable of performing the defining functions of beliefs. We will then be free to look for the defining attributes of beliefs among relationships between true beliefs and the actual world outside. False beliefs will then appear merely as things that were "supposed to" have had such and such relations to the actual world. (pp.17-18)

Dennett (1981) writes:

For the concept of belief to find application, two conditions, we have seen, must be met: (1) In general, normally, more often than not, if x believes p, p is true. (2) In general, normally, more often than not, if x avows that p, he believes p [and, by (I), p is true]. Were these conditions not met, we would not have rational, communicating systems; we would not have believers or belief-avowers. (p.19)

Alexander Bird (2007) uses the term "essential functions" and writes that belief aims at the fulfilment of its function "constitutively" (p.94).

Nolfi (2015) writes:

[B]eliefs are just the sort of attitude that is meant do a certain distinctive job in believers' mental economies. Put in other terms, what makes belief the kind of mental attitude that it is just that beliefs are supposed to fulfill a distinctive proper function. I call this account of the nature of belief the proper function privileging account. ... On the account that I am proposing here, the explanatory core of the nature of belief consists in belief being a mental attitude with this particular, distinctive proper function. (p.28)

These quotes make it clear that these authors are not only looking to functions as potentially elucidating the nature of belief, but that they expect them to demarcate what is belief from what is not belief. What are the prospects of defining belief by its function? Millikan, Bird and Nolfi draw the analogy between the way belief is defined by its function and the way hearts and kidneys are defined or constituted by their functions. But in how far are biological traits generally defined by their functions?

### §6.3.1 Functional monism

Say you think that belief is defined by its function and that the function of belief yields its norm: belief is the thing that was selected for doing x and belief does x if and only if it functions. That makes it sound as if belief has exactly one function. Does it? The teleosemanticists didn't think functional monism is obvious. They have, since the inception of their research programme, engaged with the problem of functional indeterminacy (Millikan, 1991; Neander, 1996; Price, 1998). They are worried that since we can capture a trait's function in different ways, it is unclear which of those gets to determine the content of intentional devices.<sup>67</sup> I will assume that there is an answer to the determinacy question. I will consider a different kind of pluralism, which is far less often touched upon in the literature. An exception is the following (Neander, 1995):

---

<sup>67</sup> There are actually several problems of indeterminate functions, but I won't have the space to spell them out.

[A] particular trait can have a plurality of functions. Traits can be selected for more than one effect and have more than one function without any conflict in biological norms arising. Feathers can have the function of aiding flight, regulating temperature and looking sexy to conspecifics, for example. Deformed feathers might function properly with respect to temperature regulation while malfunctioning with respect to flight, but we can simply relativise the normative judgement to the specific function. (p. 113)

It is not obvious to me whether the solution to relativise to specific functions works: what if instead of the feather we have an intentional device with multiple functions? Functions determine the content of intentional devices on teleosemantics, but if a device has two distinct functions (like the feather does), which one determines its content and thus its truth conditions? Trouble could be made for teleosemantics if intentional devices could have multiple functions. But teleosemantics is not the target of this thesis. I will focus on the relationship between functional definitions of belief and functional monism. If functional monism is false, the door is opened for significant difficulties for that brand of doxastic normativism (widely construed). So let's investigate functional monism.

There are many biological traits that have multiple functions. The wings of the Common Guillemot have flying as one function and propelling the bird underwater as another. One might say that these are one and the same function, which can be expressed generally in terms of helping the bird move. But take the example of haemoglobin, the composition and structure of which are well-understood, which plays an important role in processes sufficiently diverse to make it clear that this protein has multiple functions (Giardina et al., 1995). Haemoglobin plays a role in the metabolism of iron and has a function of signalling senescent red blood cells to macrophages, along with the function of transporting oxygen through the body. In light of that, simply assuming that belief has a single function is not good enough.

If belief had multiple functions, a classification question would open up. The definition of belief is supposed to provide us with a criterion to distinguish beliefs from non-beliefs. Say my beliefs were selected for two functions like the above, but now we find some population in which cognitions were selected only for the alethic function: are those beliefs? The Guillemot's wings were selected for swimming as well as flying, but in non-oceanic birds there is no swimming function. And what if belief had even more functions, like haemoglobin does? Say my belief-producing mechanism has five functions. Does that mean that every candidate-belief in every other organism needs to be produced by a cognitive mechanism with those same five functions? If not, where do we draw the line? How many of the functions of a given trait of mine do traits in other organisms have to share to count as the same kind of trait? There may well be good answers to these questions, but it should be clear that the picture of belief being defined by its evolutionary history is significantly complicated by the possibility of multiple functions.

### §6.3.2 Conflicting functions

The possibility of functional pluralism is especially worrisome for teleo-functional definitions of belief because of the existence of conflicting functions. Many theories have been put forward about the interaction of sexual selection functions and organism-survival functions, suggesting that in species with extreme traits that aid in sexual selection, those same traits often come at a cost to another function<sup>68</sup>. In certain cases, this has been studied in detail. The elongated caudal fin in fish of the genus *Xiphophorus* is a result of sexual selection (Basolo, 1995; Bisazza, 1993).

---

<sup>68</sup> Sexual selection is not the only reason for conflicting functions: see discussion of the toucan bill below. The thermoregulatory function of the bill was only discovered in 2009, up until then sexual selection was suspected to be the reason behind the extremely enlarged bill.

In *Xiphophorus Helligeri*, the length of this fin has been shown to adversely affect the fish's escape abilities (Basolo & Wagner, 2004). In *Xiphophorus Montezuma* it has been linked to increased oxygen expense (Basolo & Alcaraz, 2003). These fishes therefore provide examples where a single biological trait (a specific fin) fulfils its function of corresponding to female preferences at the cost of its function of locomotion. Similarly, differences in tail length due to sexual selection between the Anna's Hummingbird and the Scissor-Tail Hummingbird have been associated with differences in metabolic cost of flight of up to 11% (Clark & Dudley, 2009).

Take the bill of the toucan bird, the function of which is to help the bird eat, but also to regulate its body temperature. The latter is the reason why the bill is so large. Assume, which is plausible, that the toucan could eat more efficiently if its bill weren't so large. The large size certainly isn't advantageous while eating, as borne out by the fact that birds that do not have an additional function for their bills all have much smaller bills. How do we define the bill via its functions? Firstly, there is again the inter-species typing problem mentioned above: if something needs to have both functions to be a bill, very few other birds will count as having bills. If we scratch the temperature regulation from the definition and stipulate only the eating function as constitutive for being a bill, then that function category will lose almost all explanatory power when applied to the toucan, a major part of whose body we can only make sense of through the temperature regulation. We will therefore want to index constitutive function by species: the functions do not define the trait "having a bill" but the trait "having a toucan bill" (this already seems like a major theoretical loss to me, but let's explore the option). Hence we arrive at the definition that the toucan bill is the structure that has eating and body temperature regulation as its functions.

Now recall however that functions are supposed to fix norms of *functioning*. Being able to accommodate for this normative element of possible malfunctioning in biological and intentional devices was the whole motivation for moving away from causal functionalism and developing the notion of proper function. But when does the toucan bill malfunction? It seems that it will malfunction almost by definition: a beak that is well-suited for eating (which most toucans do not have) cannot perform the body temperature regulation task (for which it needs to be enormous), whereas a beak that better fulfils the function of temperature regulation displays some malfunction when it comes to eating. Imagine that belief had conflicting functions. This would mean that the evaluative belief norm that teleo-functionalism is supposed to generate (and thus explain why false beliefs are bad) are compromised. The norm of belief is supposed to provide us with requirements for positive normative evaluation; if we end up with two conflicting norms of belief it will turn out that there are many cases where our beliefs *cannot* fulfil their normative requirements. True beliefs would also sometimes be malfunctioning beliefs, and the explanation of the badness of false beliefs no longer works.

### §6.3.3 Conjunctive functions?

We have seen some examples of biological traits that have multiple functions, and that these can sometimes conflict with one another. If this is true of a given trait, it becomes difficult to define that trait by its function. But can't we save function-constitutivism by appealing to the conjunction of functions that a given trait has? Call this version of functionalism "conjunctive-function-constitutivism". Haemoglobin would then be the thing that exists because it transported oxygen *and* signalled cell senescence in the organism's ancestors. Situations where the jointly necessary functions conflict in such a way that they cannot both be fulfilled perfectly

will not have implications for what the function is, since functionalism doesn't require that a structure functions perfectly all the time. Normal functioning of any one structure then consist of fulfilling function a and function b both to a certain minimally required extent and in a practically achievable pareto-efficient equilibrium to one another. On the conjunctive-function picture, the examples I discussed above are not examples of traits with conflicting functions at all: instead of two conflicting functions, the fish fin just has one function that consists of a trade-off, and it is this overarching conjunctive function that defines the structure in question.

For belief norms this would still be bad news, because it could entail that our belief processes actually malfunction far less or far more often than we may have thought. Perhaps wishful thinking, for example, is not actually bad by the lights of the function of belief, but is simply a trade-off between two functions: some functionality regarding truth is conceded in order to fulfil another function belief may have (such as making the believer feel good).<sup>69</sup> That reasoning does not succeed however, because a false belief is not just a belief that performs a little less well in its truth-representing function the way an impractically-sized toucan bill is a little less adequate for eating. A false belief doesn't function at all qua truth-representer. A point that falls on the axis can still be pareto-efficient: in wishful thinking cases a belief could not have been any less false without losing some of its feel-good functionality. But this is why the definition of conjunctive function features the demand that both functions be fulfilled to a "minimally required extent". Since the truth-representing function of belief is binary, we are left

---

<sup>69</sup> Another speculative candidate for such a conflicting function could be favourable representation of the in-group and/or unfavourable representation of the out-group. Take some individual who systematically overrates the moral worth of their children, or someone who systematically underrates the moral worth of members of other cultures. The beliefs in question fail to fulfil their truth-representing function, but could instead be fulfilling conflicting functions like the in-group-favouring one.

with two options for the scenario where belief has two functions. Either, we set the minimally required functioning of truth-representing to zero, so that a belief that fulfils some other function (like making the believer feel good), counts as functioning regardless of whether it is true or false. Or, we stipulate that the truth-representing function is non-negotiable, so that all false beliefs malfunction, and true beliefs function when they additionally make the believer feel good.

The first would radically reduce the importance of falsity to the normative standing of belief. The second would reduce the importance of truth to that normative standing somewhat, and increase the importance of non-epistemic factors. I do not have the space to investigate any of this, but it seems that conflicting belief norms would present difficulty mostly for the project of grounding our norm of belief in belief's function, where there would be revisionary consequences to the content of belief norms. A less ambitious research project of simply explaining that there are *some* norms of belief that come from its function, where these are not necessarily identical to or grounding of epistemic normativity as understood throughout this thesis, can perhaps accommodate conflicting functions. That, however, would not yet solve the inter-specific typing problem above (if functions are constitutive of traits, how do we explain that the same trait can exist in multiple species but have more functions in some than in others?). Furthermore, conjunctive functions will only solve one kind of function plurality. There is another kind: diachronic function plurality, where a single trait can have several functions at different moments in time. It is here that the strongest counterexamples to function constitutivism can be found.

#### §6.3.4 Shifting functions

A single trait can vary in function across time. A well-studied example are bifunctional enzymes. These structures can perform two different tasks in the metabolism of a single organism. We often speak of primary and secondary functions, depending on which is most important or frequently used. If the environment changes so that the secondary function becomes more important, that one will become the new primary function. Now consider a case where the original primary function ceases to be called upon at all (until the environment changes again). The *Escherichia Coli* bacteria in which these transitions were studied did not evolve to the point of being a new species by any measure.<sup>70</sup> The enzyme itself, and presumably the genes that encode it, do not necessarily change much either. It seems straightforward to speak of the possession of such an enzyme with bifunctional potential as a single trait in a single species. Yet the etiological explanation for the possession of that enzyme, and thus the enzyme's etiological function, will differ between populations.

To consider a more macroscopic example, take human hairlessness. The traditional explanation (Carrier et al., 1984) for this curious trait is that it evolved to equip humans for long-distance running by enhancing thermoregulation (although there are rival hypotheses (Sutou, 2012)). Since the dawn of agriculture, that is no longer necessary. Sparse body hair has however taken on a significant cosmetic importance. If the positive cosmetic status of hairlessness persists for long enough, and if sexual selection becomes a sufficiently strong explanatory force of allele

---

<sup>70</sup> Such generalism in enzymes is theorised to have been crucial in early evolutionary times, because it has the advantage of evolvability at the cost of the effectiveness/activeness that comes with specialism (effectiveness vs evolvability are in a sense functions themselves for the meta-traits of specialism and generalism).

frequency for the trait in question across the population, it is easy to imagine a case where the etiological function of hairlessness has shifted from thermoregulation to mate appeal. A similar scenario can be imagined for bulky muscles. We have the traits that enable these now have because being muscular used to confer a direct survival benefit on our ancestors while hunting and escaping. Those environmental pressures are now becoming evolutionarily irrelevant, whilst the trait has the potential to become increasingly relevant to sexual selection. And this kind of shift in function is not unimaginable for cognitive traits. Indeed there are theories, like Miller's (2000), of the origin of human intelligence that claim intelligence is sexually selected for.

When a trait's function shifts through time, we do not speak of a new trait unless there are sufficient differences in the phenotype. Without such physiological changes, hairlessness is still hairlessness, it is still the same trait. Just like the trait that makes for production of a specific enzyme is still the same trait if *that* enzyme continues being produced. This gives us cause to think that the connection between trait and function is a loose one. The examples I now turn to demonstrate just how loose the connection really is, and just how artificial it therefore is to define traits by their etiological function: a single trait can have two literally opposed functions.

Ex. 6.1: A species of polymorphic frog has members that are red, green, or both. In the multicoloured ones, proportion of red to green differs from frog to frog, with some of them being mostly red or mostly green and others being evenly chequered. Now we take all the frogs and split them randomly into two populations. We place the first group in a rainforest with plenty of food but with lots of potential predators who all hunt using colour vision. Food availability has little adaptive pressure in the forest, but escaping predators is crucial. We place the second population of frogs in a predominantly red

environment. In this habitat, both food and predators are scarce. There are however frog-loving humans with colour vision who feed the frogs whenever they notice them. Escaping predation has little adaptive value in this environment, being noticed by humans and hence fed, has lots.

Being green is advantageous in the first environment because it means being camouflaged from predators. Being green is advantageous in the second environment because it means standing out to humans who provide food. We can realistically assume that skin colour is determined by alleles that are straightforwardly selected for. Colour tends to change very quickly under adaptive pressure, and different colour morphs that can interbreed exist for many species. Given these assumptions, the two frog populations evolve towards almost exclusively green frogs in population 1 and almost exclusively green frogs in population 2 in a short time span, without genotype changes that would warrant speaking of two distinct species. We now have two frog populations that are the same species, that share their phenotypic colour trait (they're mostly all green now), but for whom their green colour has different, indeed opposed, functions: blending in for group 1 and standing out for group 2. If biological traits were defined by their function, we would now have *different traits* in the two frog groups. But this seems absurd; they both have green skin colour. Defining traits by function seems misguided. Of course, thus far we've just looked at cases of tail length and beak size for conflicting functions and cases of hairlessness and skin colour for shifting functions. Function constitutivism could fail generally but still be true for belief. So why should the teleo-functionalist be worried about frog colour?

### §6.3.5 The functions of belief

Daniel Dennett (1981, 2009), Ruth Millikan (1984, 1986), David Papineau (2013), Peter Graham (2014b), Alexander Bird (2007) and Kate Nolfi (2015) all write about belief in biological terms. The last five also make use of the concept of function to give broadly speaking evolutionary accounts of belief. On these accounts, belief possession is conditional on a certain evolutionary history: belief is that thing that has contributed to ancestor survival through a specific function. But that means that candidate-beliefs resulting from cognitive traits that do not have the right evolutionary history are barred from counting as beliefs. So teleo-functionalism seems to involve the potential for state exclusions just like the other kinds of doxastic normativism widely construed.

Teleo-functionalism's potential for state exclusion has been noted in the literature before. (Glüer & Wikforss, 2013b) mention the Swampman problem: "An individual may behave exactly like us, like a rational agent and full-fledged language user, and yet not have beliefs simply because of its history" (p. 147). Swampman, the hypothetical creature going back to Davidson (1987) created by lightning in a swamp, is a cell-by-cell physical duplicate of a human. If intentionality or epistemic warrant are due to etiological functions, Swampman doesn't have either. Davidson levelled Swampman against the teleosemanticists, but the objection has been adapted to theories of epistemic warrant as well (Simion, 2015). The functionalists have defended their theory against Swampman. Graham (2014), for instance, writes that we must not underestimate how quickly functions can be acquired, so that Swampman won't be beliefless for long. Boyce and Moon (2016) argue that the entire appeal of the Swampman objection rests on a confused

intuition. I will not rely on Swampman for what follows, but focus on what could be the case for existing beliefs.

State exclusion based on evolutionary history is particularly problematic compared to other kinds of state exclusion we have encountered, because the teleo-functionalist criterion for belief differentiation is subject to massive ignorance. We are quite ignorant about the actual evolutionary history of real human believers, and even more so about the number of other creatures on Earth and in the universe that are endowed with representative mental states similar to our beliefs, or about what their evolutionary histories might be. In light of that, the teleo-functionalist definition might not just exclude a handful of plausible candidate-beliefs, but vast amounts of them, subject to future discovery. It could even - at its most extreme - entail that there are no beliefs, should it turn out that the assumptions about our evolutionary history that the definition is staked on are false.

To avoid state exclusion, the teleo-functionalist would have to show that it is impossible for something that is descriptively like belief to evolve in any other way than with the truth-promoting function they envisage. This kind of epistemic optimism about evolutionary history is at work throughout the literature. But it is misguided. The final cases for this chapter are meant to show why. Consider first a highly simplified evolutionary story about belief-producing traits that tend, under normal conditions, towards truth.

Ex. 6.2: A population of humans lives in an environment where material resources are the main fitness-limiting factor. They don't have to worry much about predators or enemies. Few traits are subject to pressure from sexual selection. The main challenge is finding food. These circumstances make true beliefs favourable. If one believes that the

rabbit is in the hole, that the crop will carry if one waters it, that the red berries are edible etc. in circumstances where those things are true, one is more likely to find food. Therefore, over time, humans with belief-producing mechanisms that tend to produce true beliefs have higher survival rates, thus being more likely to pass on those traits. This history means that in the present generation of these humans, belief-producing mechanisms have true representation as their etiological function.

This is the kind of story that teleo-functionalism draws on. It may well be representative of our evolutionary history, but not necessarily so. To see why, we can alter the case:

Ex. 6.3: The same original population of humans lives in the same environment except now extra-terrestrials arrive. These aliens are materially benevolent but epistemically deeply perverse. The aliens are much smarter than the humans and technologically much better equipped, and they immediately take control over all food resources. However, they like feeding the humans with the food they control, and they particularly enjoy feeding humans who have false beliefs, for they find false belief appealing in humans, much like humans find certain respiratory handicaps appealing in little dogs. The aliens are somehow very good at knowing what the humans believe. In this environment, being bad at forming true beliefs is a considerable advantage, and consequently the frequency of stupidity-enhancing traits increases. The humans also adopt a culture that promotes production of false belief. They try to make as many of their decisions as quickly as possible, ideally under time pressure, if possible whilst drunk or sleep deprived or after having hit themselves on the head. They take other maximally drunk etc. humans as their epistemic authorities. Let this perverse scenario continue

over sufficiently many generations, and we will arrive at a population of humans whose conviction-producing traits exist because they tended to produce false beliefs in their ancestors.

In Ex. 6.3, we see creatures for whom misrepresentation is the function of their candidate-beliefs. On teleo-functionalism, the pet-humans' mental states<sup>71</sup> do not count as beliefs. Furthermore, because they do not have representing truly as their function, there is no sense on which we can evaluate their falsity negatively: the false cognitions are functioning correctly, they are successful cognitions. This is not lawless deviance but something worse: deviant law. The pet-humans' false cognitions are not beliefs and are thus not censured by the norm of belief, but they are some other epistemically perverse type of mental state that fall under the perverse norms of that type of state, by the lights of which they not only escape censure but get endorsement instead. Constitutive teleo-functionalism about belief cannot be right.

#### §6.4 Against teleo-functionalism

Wedgwood (2009) mentions the worry that whatever functional account of belief we can give, there will be other mental states that fall under that description as well: functional definitions are not sufficiently precise. Whilst I did not pursue this line of argument myself in this chapter, I agree with Wedgwood that it will be very difficult to use proper functions to distinguish belief from weak affirmative attitudes. The problem of state-differentiation that we have encountered throughout this thesis could therefore be spelled out against teleo-functionalism as well. I have focused instead on the opposite worry: functional definitions for belief risk being too narrow,

---

<sup>71</sup> We can stipulate this even if the conviction-producing traits still produce more true convictions than false ones, as long as it is the past success of the false ones that explains the cognitive traits' present existence.

and certain phenomena we might want to characterise as beliefs will fall outside of the definitions' scope.

We saw that the assumptions of functional monism and function-constitution at work in teleo-functionalism are not warranted by the way functions usually work in biology, i.e. in a pluralist, highly flexible, and at times conflicting way. Without relying on synchronic functional monism, using functions to define mental states will be very complicated. Conflicting functions are especially problematic because they make complying with the evaluative belief norm difficult. Without diachronic functional monism, function-constitution is open to counterexamples like Ex. 6.1 and Ex. 6.3 and it, when applied to belief, entails the possibility of overdiscriminating state exclusions with normative consequences even worse than lawless deviance.

In the face of the plasticity that the functions of biological traits generally display, teleo-functional definitions of belief cannot assume monism and function-constitution without reference to evidence of belief's evolutionary history. Such evidence is very much outstanding. Imagine that neuroscience would reveal tomorrow that (some or all) beliefs are produced by traits that are not explained by their contribution to ancestors' true beliefs. We would not want to conclude that we have no or very few beliefs just because the underlying traits have the wrong kind of evolutionary history. We should not base our definition of belief on speculative etiology.

## Chapter 7: Towards a Knowledge-First Theory of Belief

### Doxastic involuntarism, evidence, and state-centrism

We have encountered many difficulties for normativist theories of belief. We have seen them struggle to define belief on the one hand in such a way that it is adequately distinguished from states that do not strike us as beliefs, like weak affirmative attitudes, but on the other hand so that plausible candidate-beliefs like wishful thinking are not excluded. We have seen candidate norms for belief that end up applying to all affirmative attitudes, rendering them normatively indistinguishable. We have seen that without a condition that is unique to belief to ground its norm in, the norm risks obliging subjects to hold more affirmative attitudes at once than possible, or on the other hand risks not obliging subjects to anything, making the epistemic norm an incentive for doxastic under-commitment. In this final chapter, I offer solutions to some of these problems and argue that we need to think about belief norms in a way significantly different from the current conception in order to do justice to the full range of epistemic normativity that applies to mental states.

#### §7.1 Differentiating Belief

I have classed belief as one of the affirmative attitudes. It shares an element of *confidence* with other affirmative attitudes as well as with credences, and it shares an element of *commitment* (the affirmation of its content, the representation of its content as true, having a stake in its truth) with other affirmative attitudes, but not with credences. Both confidence and commitment distinguish affirmative attitudes from attitudes like imagining or entertaining. A theory of belief

needs tools to distinguish belief from its affirmative kin. One of the main conclusions of this dissertation thus far has been the failure of doxastic normativism to provide such a distinction through normative criteria<sup>72</sup>. I now turn to the task state-differentiating task myself.

### §7.1.1 Belief and Knowledge

The relationship between belief and truth, evidence, and practical reason differs from the relationship the weaker affirmative states have with these respectively, but as we have seen not sufficiently or in a way that would yield a workable definition of belief. I have alluded in Chapter 4 to a feature of belief that seems genuinely unique: its relationship to knowledge. I said there that a version of normativism that takes knowledge as the criterion for the norm of belief (as opposed to truth or evidence) fares much better in light of many of my objections. Yet some worries remain even for knowledge variants, like the origin of obliging belief norms (see §7.4), the possibility of lawless deviance, explaining the existence of the normativity of the weaker affirmative states, and the potential counterexample of creatures capable of belief but not of knowledge. We do not have to construe belief's relationship with knowledge normatively though. Here is a perfectly descriptive fact about belief: belief is the mental state with which one can know. There are many things one can believe that one cannot know, because they are false, or blindspots, or not sufficient evidence is available to one to reach the knowledge threshold, but a thing one does not believe one cannot know by definition. This follows from one of the most uncontested facts about knowledge: that knowledge entails belief.<sup>73</sup>

---

<sup>72</sup> See especially Chapters 2 and 4.

<sup>73</sup> Although, of course, there are Radford cases (Radford, 1966). My stance on them is that their subjects do in fact believe the known proposition, but that their belief might not be occurrent or internally accessible. This kind of reply has been argued for sufficiently in the literature, e.g. by (Rose & Schaffer, 2013)

### §7.1.2 Knowledge-compatibility

Belief is the thing we know with. The question is how to make that sharp in a way that renders it useful to the venture of differentiating belief. A first step is the idea that when one believes  $p$ , whatever stands in the way of one's knowing  $p$  is not doxastic in nature: one's not knowing something one believes is not a matter of one's confidence or one's commitment. But this is still not very clear, and subject to counterexamples that take much refinement to exclude<sup>74</sup>. A closer approximation is the idea that belief is compatible with knowledge: when one believes  $p$ , one affirms  $p$  in a way that would be compatible with knowledge. We need to be careful here not to index the compatibility to propositions: it is not the case that *one's belief that  $p$*  would be compatible with knowing that  $p$ , for we can believe things we cannot possibly know. It is *the way in which* one believes  $p$  (one's confidence and commitment) that is the subject of compatibility. I propose a thesis about belief I call *knowledge-compatibility* (KC), according to which belief is that kind of affirmation with which one could know something.

Def. 7.1 knowledge-compatibility:  $S$  believes  $p$  if and only if there is a possible world in which someone knows something with at least that kind of confidence and commitment with which  $S$  believes  $p$ .

When I have an affirmative attitude towards  $p$  that is such that no one, in any circumstance, could know anything with the confidence and commitment that my attitude consists in, that attitude is not belief. Is KC true? Can there really not be beliefs that one holds so weakly that nothing could be known that way? I will offer only some brief considerations for thinking that

---

<sup>74</sup>An ingenious counterexample, due to Beau M. Mount, is the following. A subject holds the belief "I hold all my beliefs with credence 1". But her credence-introspection is slightly off, so that she actually holds all her beliefs with credence 0.95, including the belief that she holds all her beliefs with credence 1. So that specific belief cannot amount to knowledge, because it is false, where it is false because of a lack of confidence.

knowledge-compatibility holds, for it seems to flow naturally from the way in which belief has been thought about by philosophers for a long time. The age-old, albeit poorly-posed, question what we need to add to belief to get knowledge is usually asked, and attempted to answer, on the assumption that we will not have to add anything *doxastic* to the belief to make it fit for knowledge. There is evidence from outside of the field as well, from assertibility data. It seems that asserting belief whilst denying knowledge-level confidence is infelicitous. We can speak of believing whilst missing the confidence requirement for some standard:

Ex. 7.1: "I'm not so confident I'd bet my life on it, but I believe he'll be there."

Ex. 7.2: "The judge believes the defendant is guilty, but she does not feel absolutely certain."

Belief-confidence and the confidence involved in betting one's life on something or in feeling absolutely certain can be *contrasted*. This does not work however if the standard the confidence level of which we contrast belief with is knowledge:

Ex. 7.3: "I believe that I am healthy, but I am not convinced enough to know it."

Ex. 7.4: "I believe he did it, although I have too many doubts to know that."

Ex. 7.5: "I haven't made up my mind sufficiently to know this, but I believe I'll come to the party."

Examples 7.3-5 sound infelicitous. Contrast this with the way we speak about our evidence: we can say that we have evidence, but that our evidence does not suffice for some standard, along the lines of "I have some evidence, but not enough evidence". Crucially, for evidence we can speak felicitously of knowledge as the standard that is not met:

Ex. 7.6: "I have good evidence that I am healthy, but I don't have enough evidence to know it."

Ex. 7.7: "I have heard a convincing case that he did it, but the case is not strong enough to know."

What the infelicity of this way of speaking about belief suggests, is that we cannot simultaneously believe but not be confident enough to know. Indeed, belief seems defined by knowledge-level confidence in a notable way: its knowledge-compatibility is a necessary feature of it, and a differentiating feature in the sense that things that are *not* knowledge-compatible are not beliefs. Think about suspicions, guesses, being inclined to believe etc.: none of these are the kinds of attitudes with which we know.<sup>75</sup> When suspecting something true that one actually has good evidence for, one still does not know. And indeed, we can combine ascription of weak affirmative attitudes with lack of knowledge-level confidence felicitously:

Ex. 7.15: "I don't know it was him, I merely suspect it."

Ex. 7.16: "Tom knows the way" – "I never said I was that confident, I made it clear it's just an assumption it's this way."

Note that in these cases, lack of knowledge-level confidence is not just presented as *compatible* with possession of the weak affirmative attitudes, it is implied that their possession is *evidence for* the lack of knowledge. This suggests that states which are not beliefs (and usually when we ascribe suspicion and assumption we mean for them to be attitudes distinct to belief, not species of it) are incompatible with knowledge.

Something about the way in which we hold suspicions is not compatible with knowledge. You might think that there are also things about the contexts in which we form suspicions that is not

---

<sup>75</sup> Although for possible exceptions see the last paragraph of this subsection.

compatible with knowledge: we typically form this attitude on relatively slim evidence, and often about matters we have an emotional stake in that makes us less reliable. But that is a separate issue: for knowledge-compatibility the only thing required is that states that aren't beliefs are not compatible with knowledge *qua affirmation*: the confidence and commitment towards their proposition is not the kind with which one can know. The same seems plausible about guesses, conjectures, assumptions: something about them, something about their way of affirming their content, stands in the way of knowledge.

Is belief really the only kind of state that is compatible with knowledge? What about being certain? What about having committed credence 0.95? Of course, these are states a subject can be in towards a proposition she knows. But that is not a counter-example to KC. KC does not claim that belief is the only type of mental state that is knowledge-compatible, it says that any token state that is knowledge-compatible is a belief. So there can be other types of state that obey compatibility as well, as long as they are sub-kinds of belief. And this seems obviously true: if I can know something I am certain of, or if I know something with credence 0.95, then I must count as believing the proposition in question as well, for knowledge entails belief.

It's less obvious what to say about the weak affirmative attitudes. Are they really barred by knowledge-compatibility by definition? As I said above, it is plausible to think of things like suspicions as not suitable for knowledge, and some natural language practices seem in line with that theory. But it is not entirely clear that the words "assumption", "hypothesis", "suspicion",

"estimation", "conjecture" and "guess" carve out sets of attitudes all of which are not beliefs.

There is some linguistic evidence that they are usually seen as distinct<sup>76</sup> from belief:

Ex. 7.17: "Jim said the party starts at 9" – "Really? I thought he said he was guessing"

Ex. 7.18: "Do you think Jane took the money?" – "No, but I have a suspicion it was her."

But perhaps there are cases of suspicions that are beliefs. We've already noted some semantic ambiguity in some weak affirmative attitude terms, perhaps there is more, or perhaps "suspicion" and "guess" just have one meaning under which they denote a kind of attitude that has potential overlap with belief, the way belief has overlap with being certain. Be all of that as it may, KC stands. If suspicions and hypotheses are sometimes beliefs, then those token-suspicions and -hypotheses that are beliefs will be knowledge-compatible. Anything that is not knowledge-compatible is not a belief, which doesn't necessarily mean it will fall into some neat category of its own. I will continue to speak of "weak affirmative attitudes" the way in which I defined them in §2.3: those affirmative attitudes that are not beliefs. If it is true that our words for some of them, like "suspicion" can mark both token suspicions that are not beliefs and token suspicions that are a kind of belief, then only the former kind will count as a weak affirmative attitude.

---

<sup>76</sup>A paper by John Hawthorne et al. (Hawthorne, Rothschild, & Spectre, 2016) argues exactly the opposite: that belief is actually a much less committed attitude than is usually thought, and that evidence for this can be found in the infelicity of ascribing lack of belief conjunctively with possession of a weak affirmative attitude. I think there is a good pragmatic explanation available for why the conjunctions in their paper sound bad, but I cannot state that case here. Meanwhile, on their view it would be much harder to explain why Ex. 7.17 and 7.18 sound natural.

### §7.1.3 Knowledge-compatibility and conceptual analysis

With these motivations in place, let us see what follows from knowledge-compatibility. KC ties the investigation of belief to knowledge. In that sense, it falls into a knowledge-first framework: in the enterprise of trying to understand and illuminate a key concept in epistemology, knowledge gets explanatory priority. But KC does not explain belief in terms of knowledge normatively. This avoids worries about normative state exclusion (no matter how normatively deviant a certain mental state may be, it will not be disqualified from belief because of it, and will thus continue to fall under the scope of the norm of belief). It also does not fall prey to the potential counterexamples of creatures that are capable of belief but not of knowledge. Knowledge-compatibility merely demands that the way in which something is believed *would* be suitable for knowing *something*, where both the subjects of the respective believing and knowing, and the things that are believed and known, can come apart. This means that a creature none of whose attitudes ever amount to knowledge can still count as a believer, as long as the things barring her cognitions from knowledge are not doxastic.

KC would hold even if there were entire possible worlds where belief exists but not knowledge. And that may be important, in so far as the intuition that belief may be *conceptually independent* from knowledge is any good. If it is possible for belief to exist without knowledge, KC can still apply, for knowledge clearly exists in *some* possible worlds, which suffices for the right-hand-side of the biconditional to come true. But then one might think that there is something wrong with KC from the perspective of the knowledge-less world. The believers in that world won't be able to know that KC applies to their beliefs. But that's unproblematic, since they don't know anything by definition of their knowledge-less world. That creatures incapable

of knowledge should have a lesser understanding of certain concepts than we do seems plausible anyway. It's not entirely clear what the next-best way of grasping a concept is that does not involve knowledge. The knowledge-less creatures however would not even be capable of having the true belief that KC holds for their beliefs. Isn't that a problem? No, because KC was never stipulated as being part of the concept of belief. Indeed, I think it is highly unlikely that it is, for similar reasons as the ones I cited against the normativist belief concept in §4.3.1. Nor am I claiming that KC is part of the constitutive essence of belief as a mental state. Given my inclination towards thinking of mental states as natural kinds, I'm sceptical we will ever get to the full essence of belief through an analytical endeavour. What I'm claiming is that KC is necessarily true of belief, and I've already expressed opposition to the idea that necessary truths always translate into truths about constitutive essences. Perhaps future research will reveal that KC is more than just a necessary truth about belief, and that might then have consequences for the possibility of worlds with belief but not knowledge. For now, I am merely stipulating knowledge-compatibility as a principle that necessarily applies to belief, and that can help us explain some of the psychological and normative facts about belief. That is the task for the next section.

## §7.2 Knowledge-compatibility and the psychology of belief

In earlier chapters, we encountered several desiderata for a theory of belief and for the formulation of its norm. Firstly, that it should provide us with a criterion for *state-differentiation*: what makes belief different from other mental states? Secondly, that it can account for descriptive explananda of *evidential sensitivity*, which includes explaining the differences between belief and other mental states regarding these psychological phenomena. Thirdly, that

it should be able to account for the *normative explananda*: the force and authority of the norm of belief, truth and evidence as normative criteria, and the normative similarities and differences between belief and the other affirmative states along both the force as well as the criterion dimensions. Finally, that it can avoid some of the problems we have identified for doxastic normativism: the *over- and under-prescription dilemma* and the *discrimination problem* for belief's distinguishing criterion.

### §7.2.1 Doxastic involuntarism

Many defences of normativist accounts of belief take the supposed *involuntariness* of belief as a premise. Involuntarism is the thesis that we do not have voluntary control over our beliefs, often captured as the inability to "believe at will". The literature often distinguishes between indirect doxastic control and direct doxastic control, where the former is commonly accepted and the latter is commonly rejected (Pojman, 1999), often even declared conceptually impossible (Williams, 1970) (Scott-Kakures, 1994). The perceived impossibility of believing at will has been linked to ideas about belief's intrinsic normative features at least since Bernard Williams (1973). We have studied a version of an argument that takes an involuntarist premise to establish conceptual normativism in Chapter 4. As we saw there, formulating that involuntarist premise is difficult, as is capturing the exact difference between belief and other affirmative attitudes when it comes to questions of voluntarism.

I do not take a stance on the truth of involuntarism in this thesis. This may ultimately well come down to an empirical question, and our study of belief may well be too young to yield a definitive answer at this stage. But some observations can be made. One of them is that much of the support for doxastic involuntarism seems to stem from the observed difficulty of

controlling one specific kind of doxastic event: the formation of belief from a starting point of neutrality (as opposed to from a starting point of a weak affirmative attitude). And it is true that getting oneself to believe that *p* at will from a situation where one has no attitude towards *p* whatsoever seems very hard. Let's for the sake of the argument grant that it is not only hard but indeed impossible. Belief formation, however, is not the only kind of doxastic event: belief abandonment (going from belief to no belief), which can come in the form of suspension or in the form of extinction through a rival belief, upholding (going from belief to belief), and withholding (going from no belief to no belief) are just as important in bringing about epistemic agents' complete belief state (everything they believe at a given point in time).

Attention is increasingly (Nottelmann, 2007) being paid to the potential differences in freedom regarding these types of events, but the focus is still disproportionately on formation. Yet evidence of the difficulty or even impossibility of forming a belief at will does not suffice to show that we lack control over our existing beliefs via deliberate suspension or deliberate upholding. I distinguish freedom to believe from freedom from belief. Freedom to believe is the freedom to deliberately form a belief one did not have before. Freedom from belief is the freedom to abandon a belief one had before deliberately. The involuntarist has a much more difficult task in arguing for the impossibility of the latter. Firstly, because freedom from belief seems well-respected in the philosophical literature where suspension of belief is generally assumed to be possible. Secondly, because asymmetry between freedom to and freedom from is not unusual among mental states. Many people for instance strive for control over their fears by learning to stop fearing something and no one thinks that kind of control over fear is compromised by the

inability to start fearing at will. Hans Rott (2017) calls the thesis that there is an asymmetry in control over belief formation and belief abandonment “negative doxastic voluntarism”.<sup>77</sup>

### §7.2.2 Kinds of freedom

I think something like negative voluntarism is likely to be true of belief<sup>78</sup>. To the freedom from/freedom to distinction, we should also add distinctions about what kind of mental state the subject is in when attempting to exercise the doxastic control (going from neutrality to belief vs going from assumption to belief for instance) and what her evidential situation is (does she perceive evidence for p, no evidence, or evidence against p?). With those distinctions in place we get a multitude of kinds of doxastic voluntarism:

Def. 7.2 very strong freedom to believe: it is possible to wilfully form the belief that p despite perceived evidence<sup>79</sup> against p.

Def. 7.3 strong freedom to believe: it is possible to wilfully go from neutrality to the belief that p despite perceiving no evidence<sup>80</sup> for p.

Def. 7.4 strong freedom from belief: it is possible to wilfully go from belief that p to neutrality despite perceiving evidence for p.

Def. 7.5 weak freedom to believe: it is possible to wilfully go from a weak affirmative attitude that p to belief despite perceiving no additional evidence for p.

Def. 7.6 weak freedom from belief: it is possible to wilfully go from belief that p to a weak affirmative attitude that p despite perceiving the same amount of evidence for p.

---

<sup>77</sup> “(i) humans tend to acquire beliefs more or less automatically and unreflectively, but (ii) they tend to withdraw beliefs in a controlled and deliberate way” (p. 2695).

<sup>78</sup> Note that as stated negative voluntarism is compatible with doxastic voluntarism about formation being psychologically possible, as long as it happens more rarely than free withdrawal of belief.

<sup>79</sup> This is about the subject’s total evidence: considering all evidence at her disposition, does she perceive it to be pointing to the falsity of p?

<sup>80</sup> No evidence for p on one’s total evidence can mean either that there just is no evidence either way, or that one’s evidence is balanced perfectly so that one’s total evidence points in neither direction.

Def. 7.7 very weak freedom to believe: it is possible to wilfully increase one's confidence or commitment levels of one's belief that  $p$  without perceiving additional evidence for  $p$ .

Def. 7.8 very weak freedom from belief: it is possible to wilfully decrease one's confidence or commitment levels of one's belief that  $p$  without moving across the threshold to another kind of attitude, without perceiving less evidence for  $p$ .

It should be clear that some of these could be true while others false. There is no reason to expect "control over our belief" to be a monolithic ability expressing itself in equal amounts of freedom over the causing or preventing of doxastic events that differ in nature. The authors who use doxastic involuntarism as a datum in their theories of the normativity of belief (e.g. Audi, 2001; Engel, 2002, 2009; Owens, 2003; Shah & Velleman, 2005) are unlikely to agree with me that very strong freedom to believe and strong freedom to believe might be possible. I do have to grant the involuntarists that possible or impossible, both of these at least occur very rarely. Let's assume for the sake of argument that they are both impossible. Can knowledge-compatibility help us explain either impossibility or extreme rareness of acts of strong or very strong freedom to believe? I will argue that it can.

But first we need to think about the other affirmative attitudes. We've seen argued in several places that belief differs from weak affirmative attitudes in its evidential sensitivity. Let's take a look at some comparisons.

Firstly, it seems likely that in so far as we do not have very strong freedom to believe, we do not have very strong freedom to form other affirmative attitudes either. I cannot guess that the present is hidden under the bed while perceiving only evidence for it being hidden inside the wardrobe. I cannot assume that there is wine in the glass when I'm looking at it and seeing clearly that it is empty.

Secondly, there are at least some affirmative attitudes for which we seem to have strong freedom to form them, more so than we do for belief. If you show me two upside-down bowls and say one of them has the biscuits underneath them, I can make a guess which one it is. If I have to run to make the train and I don't remember whether it leaves from platform 4 or 6, I can assume that it's 4 and run towards that platform.

Thirdly, we often have strong freedom from belief: there is a level of perceived evidence that is compatible with belief but also with a weaker affirmative attitude or with suspension of belief, so that we can decide which state to adopt. There are some affirmative attitudes for which strong freedom from seems rarer than for belief: deciding to stop suspecting something or stop assuming it, in the absence of a change in perceived evidence, and in the absence of a new suspicion or assumption to extinguish the existing one, is harder than it is to decide to stop believing something in most cases. For belief too though, it is possible that there is a level of perceived evidence that is compatible only with belief: we do not always have strong freedom from belief.

Fourthly, very weak freedom to and from seem likely to obtain for both belief and other affirmative attitudes, although for very weak freedom from there is again a sense in which lowering the confidence of one's suspicion or assumption somewhat without epistemic cause might be harder than lowering the confidence of one's belief somewhat.

These are four psychological explananda of evidential sensitivity that I will attempt to elucidate. Doxastic normativism fails in this task because it does not have an adequate criterion for state-differentiation. Let us look at how the knowledge-compatibility definition goes about accounting for these differences in evidential sensitivity.

### §7.2.3 Ignorance as failure

Knowledge is normatively significant. Its normative significance<sup>81</sup> should be uncontroversial regardless of one's views on the nature of epistemic normativity. But here are some specific versions of views that specify knowledge's normative significance.

1. Knowledge is intrinsically valuable
2. Knowledge is the goal of inquirers: every epistemic agent undertaking an inquiry has as a matter of fact as her goal to attain knowledge.
3. Knowledge is the goal of inquiry: every inquiry has as its goal the attainment of knowledge, regardless of whether the inquirer subscribes to that goal herself.
4. Knowledge is an obligation: every epistemic agent has an obligation to obtain knowledge.
5. Knowledge is instrumentally valuable: most actual cases of human knowledge make the knower in some sense better off.
6. Knowledge is necessary for acceptable human behaviour: since it makes for the permissibility of things like assertions and use of premises in practical reasoning, and those are things most humans want to undertake, they need knowledge.

At least some of these should be plausible normative principles about knowledge. Much of the work in epistemology throughout the millennia has borne witness to our commitment to these kinds of ideas. They fall under three kinds of epistemic principle: that knowledge is a value, that knowledge is an aim, and that knowledge is an obligation. I will assume in what follows that at least one of these three is true of knowledge, and that that normative property is expressed

---

<sup>81</sup> Mark Schroeder (2015) gives the example of death as a normatively significant concept. Death itself is presumably a completely descriptive thing and we need no norms to explain what it is. However, perhaps contingently, death is of incredible normative significance to us: "The fact that an action would result in one's own death is at least nearly always some significant reason not to perform that action. When moral philosophers are interested in the normative question of what people ought to do, their answers are rarely uninformed by assumptions about who would die on each course of action."

through something like one of the six statements above. Any of them can do the kind of explanatory work I will use the normative status of knowledge for. I will therefore refer to the disjunction of one-six as “the epistemic norm”.

The epistemic norm gives epistemic agents a reason for wanting to know. That humans, in most ordinary contexts, have a reason for wanting to know (which shouldn't be controversial), suffices for the explanation I have in mind. I call this the desirability thesis: knowledge is *prima facie* desirable. Notice that none of this is formulated as a principle about belief: knowledge is desirable for every epistemic agent at any given point in time, not because they are in a specific normative state, but because they have knowledge as a goal themselves. The desirability of knowledge that follows from the epistemic norm (and consequently the incentive to strive for knowledge) is constantly active in the background of human activities, because it is an incentive for the *agent*, not for the state of belief. Therefore, I call the epistemic norm an *agent-centric* norm. The norm of belief we see on doxastic normativism on the other hand, is a *state-centric* norm. It's a norm that applies to believers in virtue of being in a specific state. This difference will become relevant when we compare normativism with knowledge-compatibility.

What does the epistemic norm have to do with belief? Believers are epistemic agents and epistemic agents fall under the epistemic norm and hence have a reason for wanting to know. Because belief is the knowledge-compatible affirmative attitude, agents can only discharge the epistemic norm (fulfil their obligations, obtain what is valuable, reach their goal) through beliefs. This automatically puts normative restraints on epistemic agents regarding what they should do with their beliefs. A belief the content of which the believer does not know *fails* in an epistemically relevant sense. It does not help the believer discharge the epistemic norm. Indeed,

it may take her further away from discharging it than if she had not had the belief. So we're seeing a normative picture on which unknown beliefs fail to help the believer discharge a norm that she is subjected to independently of the belief but which she cannot fulfil in any other way than through belief.

For my explanation of the evidential sensitivity of belief I will rely on the following phenomenon: most believers do not like realising that a belief of theirs is unknown. On my view, the reason for this is that ignorance-ascription to an existing belief amounts to an admission of failure. Of course, the dislike in question can be very mild. One can be very used to making mistakes. Also, the ignorance-discovery can be accompanied by relief, if the content of the belief in question is an undesirable one. However, even when I am relieved to find I do not actually know some unpleasant proposition I believe to be true, there is a sense in which I dislike having believed it. A second phenomenon is that ignorance-ascription tends to lower confidence. If I believe  $p$  at  $t_1$  and form the second-order belief at  $t_2$  that I do not know  $p$ , my confidence in  $p$  tends to go down, often to the point of extinguishing the belief. Importantly, the realisation that one does not know some proposition one guesses or estimates does not tend to lower one's confidence in the guess or estimate. Also, believing  $p$  and coming to the realisation that one does not have perfect evidence for  $p$  does not lower confidence in  $p$  to the same extent that the ignorance-realisation does. The hypothesis that unknown beliefs are failures explains all of this well: not being based on perfect evidence is not necessarily a failure for a belief. While having infallible warrant might be beneficial in some contexts, there is no general obligation on epistemic agents to achieve it. Even more clearly, not amounting to knowledge is not a failure for a guess or estimate or suspicion, for these states typically don't amount to knowledge, and have a different kind of investigative function.

When a typical epistemic agent self-ascribes ignorance to a pre-existing belief without thereby weakening the belief in question so that the belief that  $p$  and the belief that she does not know that  $p$  are maintained simultaneously, the believer finds herself in the situation that is described in the epistemic version of a Moorean conjunction: "P and I do not know that  $p$ ". To many believers, self-ascription of a Moorean conjunction feels odd. I will not take a stance here on the nature of this oddness. It does not matter for the explanation I have in mind whether there is something genuinely paradoxical to these sentences, deserving of philosophical explanation as some philosophers (Douven, 2009; Smithies, 2012a; J. N. Williams, 2007), have thought, or whether believing an epistemic Moore conjunction is unpleasant simply in the way in which self-ascribing akrasia is unpleasant to most agents. What matters is that most believers do not like to consciously maintain the state described by the Moorean conjunction. This observation serves as an explanans rather than an explanandum in what follows.

#### §7.2.4 Belief and dispositional ignorance-ascription

In the previous subsection we concluded that we tend not to maintain beliefs whose content we self-ascribe ignorance for. What does that mean for evidential sensitivity? Let's capture perceived evidence in epistemic probability for now, where 0 means the subject perceives evidence that make her judge  $p$  to have no probability whatsoever, 0.5 means she perceives evidence that makes her judge  $p$  and  $\neg p$  are equally likely, and 1 means she perceives  $p$  to be undeniably true. There is a certain perceived evidence level that suffices (psychologically, not normatively) to form a belief. On strong freedom to believe, that might be 0.5, on very strong freedom to believe it might be even lower. Since we are conservatively assuming with many of the authors in the debate at hand that we do not have strong freedom to believe, the perceived

evidence level necessary to allow one to form a belief should presumably be at least 0.6 or so. For a given subject, there is also a perceived evidence level that that subject thinks suffices epistemically for knowledge<sup>82</sup>. So say some subject considers 0.85 this minimal epistemic requirement for knowledge, whereas she is psychologically capable of forming beliefs from 0.7 upwards. Now this subject forms a belief *p* on perceived evidence 0.75. This belief is *dispositionally Moorean*. If our subject were to reflect on her epistemic position, she would come to believe the Moorean conjunction that she believes *p* but does not know *p*. And to avoid being in that unpleasant state, she would probably lower her confidence in *p* below the threshold for belief.

Consider next the following postulate: in most subjects in most situations, the mental states they adopt tend not to differ much from the mental states they are disposed to adopt upon reflection. Call this postulate *minimal discrepancy*. There are obvious and significant counterexamples to this: many biases and illusions make us adopt mental states that would not survive critical self-reflection. Pathologies like phobias make us adopt mental states that do not even cease after actual scrutiny reveals them to be irrational. But although significant exceptions, these do seem like exceptions. The feelings of surprise, unease, and confusion we experience when we encounter these kinds of inconsistencies (beliefs we come to realise we don't take ourselves entitled to, fears we realise we have no reason to have) bears witness perhaps to their exceptional nature. We're not going through life constantly discovering our mental states to be out of line with how we think they ought to be.

---

<sup>82</sup> This is put in very intellectual terms. I will consider further down how the evidential sensitivity through knowledge-compatibility schema would work out for believers who may not have a concept of knowledge or be capable of reasoning about evidence.

Minimal discrepancy would make sense from a mental economy perspective: revising mental states costs cognitive energy, so the mind has an incentive to install dispositionally stable attitudes from the get go. And as epistemic agents encounter discrepancies between what they believe and what they are disposed upon reflection to believe, they learn from those experiences and their tendencies towards minimal discrepancy become finetuned. Something like minimal discrepancy would explain why the perceived evidence level on which subjects tend to form beliefs is not very different from the evidence level they consider necessary for knowledge. For our subject above, beliefs on 0.75 are dispositionally unstable: their persistence is contingent on the believer not reflecting upon her epistemic position, for if she were to form an epistemic judgment she would start believing that she does not know the thing she believes, which would be likely to affect her first-order confidence and sink it below the threshold for belief.<sup>83</sup>

Recall that evidential sensitivity of belief seems to have no hard boundaries, since the amount of evidence necessary to form a belief differs from subject to subject and within subjects from situation to situation. The explanation we considered is coherent with and indeed predictive of this kind of variability: subjects *tend* to lower first-order credence in reaction to ignorance-ascription, but they can fail to, and subjects *tend* to have mental states that are compatible with the mental states they are disposed to have upon self-reflection, but they can fail to. The picture we just considered, of subjects having a tendency towards beliefs that are like the beliefs they would have after epistemic self-reflection, explains the evidential sensitivity of belief quite

---

<sup>83</sup> Some people intuit that subjects who subscribe to demanding epistemic principles are likely to form certain beliefs less quickly or have higher need for closure. If this is true, the idea that our beliefs tend to be in line with what we would judge ourselves to know on reflection would explain it well.

straightforwardly: the lower the amount of perceived evidence a subject would attempt to form a belief on, the higher the discrepancy with what she would believe after epistemic self-reflection. It also explains with ease why beliefs are more sensitive to evidence than weaker affirmative attitudes. Guesses, suspicions etc. (except in those potential cases where they are beliefs) never amount to knowledge. There is no Moorean oddness in self-ascribing a suspicion one does not know the content of, or in saying one does not know something one assumes. Self-ascribing ignorance to such a weak affirmative attitude does not amount to an admission of failure about that attitude, for it is clear anyway that one does not know. There is thus no reason for one's weak affirmative attitudes to be informed by one's subjective epistemic threshold (the amount of evidence one deems necessary for knowledge).

#### §7.2.5 Freedom from affirmative attitudes

Thus far, I have offered an explanation of the limits to or difficulty of forming beliefs at will, and of the asymmetry with the greater ease with which we form weak affirmative attitudes deliberately on low evidence. Let's turn next to freedom from belief. I wrote as the explanandum of freedom from affirmative attitudes that we often have strong freedom from belief, meaning we can decide whether to have a belief or have a weaker attitude (including agnosticism) instead. Furthermore, there are some affirmative attitudes that we seem less free from than from belief: deciding to stop suspecting or assuming can be harder than deciding to stop believing. However, for belief too, there seems to be a level of evidence that does not allow an attitude other than belief. Can we make sense of these phenomena through knowledge-compatibility?

There are situations where our perceived evidence is so that it suffices for us to believe, but where we have a choice between believing and refraining from believing or continuing to believe and suspending belief respectively. Since the epistemic norm on my account is situated on the level of the agent instead of on the level of individual mental states, it makes sense that we can opt out of some beliefs that would have amounted to knowledge: agents can decide for specific beliefs that it's not worth risking falsehood, or that too much depends on them, or that agnosticism serves their wider epistemic goals better. If on the other hand, the epistemic norm originated on the mental state level, every single belief that would meet it would be equally obligatory, and the fact that we can make decisions about what (belief-allowing) evidence we let ourselves be made to believe by and what (belief-allowing) evidence we do not would be harder to make sense of. The agent-centrism of the epistemic norm explains why agents have leeway over which beliefs to keep and which to abandon: not every belief we have sufficient evidence for will fulfil the same role in helping us discharge the epistemic norm. Of course, that doesn't mean that our acts of suspension will always be due to wholesome epistemic reasons like deeming agnosticism a good investigative strategy for some specific inquiry. Often, we instead suspend for emotional or moral reasons. But our *ability* to suspend makes sense epistemically speaking on the agent-centrist epistemic norm. And so does the asymmetry between freedom to believe and freedom from belief: beliefs that are deliberately formed on insufficient evidence are very unlikely to help one discharge the epistemic norm, and are indeed likely to make one *less* compliant with it. The set of beliefs that are formed on sufficient evidence is much more normatively diverse: some will amount to knowledge and some will not, of the ones that do some will have more epistemic import than others (because more follows from them or more is staked on them).

The above explains why we have freedom from belief. Let's next explore how it works. Consider the phenomenology of suspending a belief one has good evidence for. In cases of suspension, we don't say to ourselves "I have enough evidence to know that he's guilty, but I'm not going to believe it". Self-ascription of being in a position to know is not very compatible with first-order suspension of belief, much as self-ignorance ascription doesn't sit well with first-order belief. Instead, we tend to focus on our doubts and say to ourselves "I don't know that". Thinking "I don't know that it's going to stay dry" usually suffices to make me not believe that it is going to stay dry. While freedom from beliefs we epistemically second-order endorse is very difficult to attain, we can easily refrain from believing on good evidence by not conceiving of ourselves as knowing. KC in conjunction with the desirability of knowledge thus accounts for our freedom from belief.

Finally, there is the difference in freedom from belief with the weaker affirmative attitudes to think about: once held, it can be more difficult to stop suspecting, guessing, conjecturing or assuming than it would be to suspend belief. This might seem counterintuitive: don't guesses and assumptions come as easily as they go? In a sense they do, because they are easily replaced with *another* guess or assumption or estimate. If one's epistemic situation allows one to guess an incompatible alternative to  $p$ , the ease with which it is possible to make that new guess (the amount of freedom to) is what explains the short-lived nature of the previous one. In the absence of a new affirmative attitude to take their place though, the weak affirmative attitudes are actually more resilient than beliefs are: if I believe the butler did it, all I have to do to stop

believing this is tell myself I don't actually know<sup>84</sup>. If I suspect the butler did it however, realising I'm not sure won't change that suspicion at all. Similarly, if I make a guess that p I'll need a better reason against p to suspend that guess than I would to suspend the belief. All of this is again well explained by belief being the knowledge-compatible attitude: the epistemic self-realizations that we can use to suspend our beliefs do not work on the weaker affirmative attitudes because these are not knowledge-compatible. It thus makes perfect sense that telling myself I don't know that p will have no influence on my possession of attitudes that are not candidates for knowledge anyway.

### §7.3 Normative differences between belief and other affirmative states

Having sketched an account of evidential sensitivity through knowledge-compatibility, I now turn to the second set of explananda normativism struggled with: capturing the right kind of epistemic normativity we expect for beliefs. We want our theory of belief be able to explain the following normative intuitions:

1. Beliefs can be true or false and false ones are mistakes.
2. Other affirmative attitudes can be true or false and false ones are mistakes.
3. Beliefs can be well or poorly warranted, and well warranted ones are better than poorly warranted ones.
4. Other affirmative attitudes can be well or poorly warranted, and well warranted ones are better than poorly warranted ones.

---

<sup>84</sup> It is worth pointing out here that when we stop believing, we don't automatically fall into one of the weaker affirmative attitudes. They are not residual beliefs and they reduce to credences as little as beliefs do. Many of them have to be entered into quite deliberately. So while it is possible to stop believing that p and then find oneself with the suspicion that p instead, this is not automatically the case when one stops believing without adopting neutral credence instead.

5. For beliefs, it is more important that they are good than it is for affirmative attitudes.

As we saw, normativism has difficulty accounting for all of them conjunctively. If the source of epistemic normativity is a property only belief has, 2 and 4 are mysterious. If it is a property belief shares with the others, 5 is difficult to account for. Let's see if KC fares better.

For belief, 1 and 3 are easy to explain: knowing  $p$  is better than being ignorant of  $p$ , and believing  $p$  is the only way to achieve knowledge of it, and false beliefs as well as poorly warranted beliefs fail to be knowledge, hence both of them are failures of the agent. But what about the weak affirmative attitudes? These do not amount to knowledge by definition, so why is it worse to hold ones that are false or poorly warranted? To explain this, it is important to draw attention again to the agent-centric nature of the knowledge norm I have in mind. Individual mental states get evaluated in terms of their contribution to the agent's overall compliance to the epistemic norm. As we have seen in §4.4.4, weak affirmative attitudes have an investigative function. They allow us to get closer to knowledge by representing the world in a more guarded way than beliefs do, enabling us to treat them as premises in our inquiries, motivating us to begin or maintain certain investigations without staking too much on their truth. We are for instance much less likely to act on weak affirmative attitudes.

Nevertheless, having false weak affirmative attitudes can be epistemically disadvantageous to us. Firstly, they can cost us time: if I guess that the kitchen is at the end of the corridor on the left, go to check to find that I was wrong, I can go back and check the corridor on the right instead, but it would have been better for me if I had guessed correctly the first time. Consequently, misguided affirmative attitudes can make us lose out on vital information. If I suspect the butler did it and then spend a week shadowing the butler to find nothing

incriminating, I lose out on the information I could have been gathering had I suspected the cook instead. Secondly, we are not infallible introspectors about the nature of our mental states. I may form a suspicion that  $p$  but then mistake it for a belief and start treating it as a belief (at which point it is likely to become a belief). Even without such introspection mistakes, it is possible that assumptions that have been held for a long time eventually become solidified as beliefs, and their hedged origin is forgotten. So while false weak affirmative attitudes involve no direct failure to know, they come with the risk of causing false beliefs, and they come with the opportunity cost of true attitudes that could have been held instead, and the investigative payoffs those alternative attitudes could have effected.

This explains how the general epistemic norm influences the normative status of affirmative attitudes other than belief. It also explains why the normative status of weak affirmative attitudes seems less important to a subject's overall epistemic standing (false guesses and suspicions are bad, but they are far less bad than false beliefs, because a false belief will be mistaken for knowledge). Finally, it explains why warrant is of much less normative concern for weak affirmative attitudes than it is for belief, and for true ones it is of barely any concern whatsoever. Luck is not a cause for complaint for guesses to the same extent that it is for beliefs. Forming affirmative attitudes on warranted grounds minimises the chances of their falsehood just like warrant for belief does, and unwarranted weak affirmative attitudes are consequently more likely to be epistemically disadvantageous due to falsehood. There might also be a certain risk of a habituation effect caused by repeated formation of weak affirmative attitudes on very slim warrant, leading one to more epistemic recklessness in the future, which might eventually affect belief formation too. But mostly, weak warrant for true affirmative attitudes has little negative effect on the subject's ability to meet her overall epistemic goals and is consequently not

something we tend to censure. Consider the rebuke “but you don’t know that!” to an assertion. If I made an explicit belief-report, or a flat-out assertion which tends to be received as a belief report, my interlocutor saying they don’t consider me knowledgeable of what I just said is serious criticism of my epistemic standing, and I will need to defend myself somehow. If however, I reported on a guess or assumption and my interlocutor says “but you don’t know that!”, I can simply respond with “obviously I don’t know, that’s why I’m making a guess”. Guarded commitment to the truth of a proposition in circumstances where I’m not in a position to know that proposition is exactly what states like guessing and assuming are for, so complaints about their limited evidential basis are out of place. The knowledge-incompatibility of weak affirmative states explains this perfectly.

#### §7.4 The failures of state-centrism

Having looked at the basic framework of the normativity of affirmative attitudes and the source of that normativity, I now turn to some problems I identified for doxastic normativism throughout the thesis, as well as some novel ones, in order to show how a move away from state-centrism can solve them.

##### §7.4.1 Over- and underprescription

It’s not obvious how to explain obliging belief norms on the normativist picture. On state normativism, beliefs have conditions inscribed in them that make them correct or incorrect in certain circumstances. A false belief violates its own condition of permissible use. But where does the normativity of the absence of a belief that ought to be there come from? It is implausible that absence of belief has intrinsic norms as well. It is a challenge the normativist needs to tackle explicitly, how to explain the epistemic fault in the absence of a belief on a

model where epistemic normativity comes from belief states. On conceptual normativism, the epistemic norm comes from reasoning about belief, which involves thinking about a normatively constituted concept. But again, what about a situation where a subject fails to form a certain belief, and fails to ask herself whether she should believe that thing? How does a norm that derives from a concept (that is not being entertained in the given situation) censure that situation? On teleologism finally, the aim of belief is formulated such that anyone who believes  $p$  thereby aims for that belief to be true. But what about when one does not believe  $p$ ? Again, it seems unlikely that absence of belief is an aim-directed state. The game analogy that is so often used to explain constitutive normativity can illustrate the problem well. Chess is constituted by the rules that govern it: something is a chess move only if it is governed by those rules. Moving chess pieces around on a board without minimal adherence to these rules or at least the intention to adhere to them is not playing chess. The rules censure certain moves as illegal, and they classify certain moves as check or check mate. But one thing the rules of chess never specify is whether one ought to play chess at all. They only start applying once one is already playing chess.

Of course the doxastic normativists, most of whom are interested in truth norms, might not find the difficulty of explaining the origin of obliging belief norms very troublesome, for truth norms have problematic consequences anyway when they are spelled out in an obliging logical form (Bykvist & Hattiangadi, 2007). As a response to Bykvist and Hattiangadi's blindspot objection, some truth norm defenders have moved away from obliging belief norms to purely restrictive ones (Raleigh, 2013; Whiting, 2010), some of which (Raleigh, 2013) within an explicitly normativist framework. But as I argued in §4.3.5, we need obliging belief norms to avoid underprescription. With merely a falsity-restricting and/or truth-permitting belief norm,

consistently under-committing agents will be ranked as epistemically better than ones that take risks in the pursuit of knowledge, for they are at much less risk of violating. Agents who systematically undercommit themselves by adopting weaker affirmative attitudes than belief will be ranked as epistemically better by the norm of belief, for they are at no risk of violating the restrictive formulation. Someone with 1000 beliefs 0.5% of which are false will do much better on such a norm than someone with 100000 beliefs 0.1% of which are false. Agents who suspend belief on questions they find unpleasant will get no epistemic censure. Agents who tend to form beliefs mostly about easy-to-know matters will be ranked epistemically better than those who also venture into more intellectually challenging investigative domains. We will have no way to censure those who refuse to adopt inconvenient beliefs they have strong evidence for. Generally speaking, without an obliging side to the belief norm, it seems that suspension of belief is the best<sup>85</sup> attitude to take on about as many matters as possible, for it comes at no risk of falsehood. So the doxastic normativist should provide a belief norm that is obliging as well as restricting.

But not only does it take a special explanation for how obliging norms are generated on the constitutive norm model, there is overprescription to worry about next: if the criterion that gives beliefs their normative status is one that it shares with other affirmative attitudes (and crucially truth and warrant both are shared criteria in this way), this will make multiple distinct affirmative attitudes obligatory at once, and while it may well be possible to hold more than one distinct attitude towards the same proposition, it seems highly unlikely that we are psychologically capable of guessing and suspecting and estimating and conjecturing and assuming and

---

<sup>85</sup> By the lights of the belief norm.

believing one and the same proposition at a particular moment in time. The doxastic normativist needs to find a distinguishing reason that explains why one can have an obligation to form some specific true belief, but not simultaneously have the obligation to have all the weak affirmative attitudes as well. As we've seen, this is a serious challenge.

The agent-centric epistemic imperative however, combined with the definition of belief as the knowledge-compatible affirmative attitude, can account for obliging belief norms without falling prey to the under- and overprescription dilemma. Since epistemic agents' ultimate epistemic obligation is to seek knowledge, and since only belief is knowledge-compatible, where belief is permissible, adopting a weaker attitude instead will never be ranked epistemically better, avoiding underprescription.<sup>86</sup> The distinguishing criterion of knowledge-compatibility simultaneously makes it so that no other attitude will be classed obligatory along with belief in any particular situation. The account can also explain why one sometimes ought to adopt a weak affirmative attitude instead of suspend belief: if the former helps one in one's investigative endeavours, it is preferable to suspension, even though neither of them is itself knowledge-apt. Similarly, it is easy to see why in other scenarios only suspension is acceptable: if a weak affirmative attitude would lead one epistemically astray by making it less likely that one attains knowledge of surrounding matters or by endangering the confidence with which one holds existing known beliefs, it is better to remain agnostic instead.

The source of epistemic normativity on my view is top-down instead of bottom-up: it does not come from the individual mental states which ought to be employed in certain ways but is an

---

<sup>86</sup> If the epistemic norm is an obligation to know, this is straightforward. If it stipulates an intrinsic value, then agents have a prima facie reason to obtain knowledge. If it stipulates an aim for token or type inquiry, agents will have failed to fulfil their aim when they don't know.

overarching goal that lends mental states normative status depending on the role they play in fulfilling that goal. This explains why overprescription won't arise among the weak affirmative attitudes (for which knowledge-compatibility offers no criteria of state distinction because they are all knowledge-incompatible): even in a scenario where multiple distinct affirmative attitudes would contribute to the agent's fulfilment of her epistemic obligations, they will not be jointly obligatory.

#### §7.4.2 State-overarching epistemic normativity

A similar problem to over- and underprescription is the need for the epistemic norm to be able to recommend some attitudes over others. There are situations where one ought to estimate instead of guess, believe instead of speculate, assume instead of conjecture, but not because there would be something wrong with the guess, speculation, or conjecture *by their lights*. I could hold a perfectly fine guess judged by some local norm for guessing<sup>87</sup>, but still be at fault for not having estimated instead. In other words, we can incur epistemic censure for the opportunity cost of our attitudes. This intuition speaks strongly against a state-centrist construal of epistemic normativity, for state-centrist norms, like those offered by doxastic normativism, can only ever evaluate mental states by virtue of that type of state's permissibility criterion. Again, the analogy with chess is helpful to illustrate the problem. Say there is a game of checkers on my right and a game of chess on my left. I could make a good checkers move on the right and a good chess move on the left, both considered by the lights of the rules of checkers and chess respectively. But for some reason I can only either make the checkers move or the chess

---

<sup>87</sup> All of this is under the assumption that the state-centrist would even be able to formulate specific norms for all the different affirmative attitudes, which is obviously a tall order.

move, not both. The rules of chess and checkers will be completely useless in helping me decide which move to make.

Because state-centrism evaluates mental states by their intrinsic properties, and because opportunity costs that make reference to which attitude could have been adopted instead are not intrinsic properties of a mental state, state-centrism cannot issue the kinds of state-overarching recommendations in these cases that our normative intuitions lead us to expect. Again, agent-centric normativity offers a solution to the problem of overarching epistemic normativity: since the normative status of mental states depends on their contribution to the agent discharging the epistemic norm, the normative status of individual attitudes will take opportunity costs into account and whichever mental state does best in a particular situation will be recommended, or the disjunction of states that do equally well. This result is ensured by the fact that trying to adopt an impossible (or perhaps merely psychologically extremely taxing) combination of mental states will almost by definition not serve the agent well in her pursuit of discharging the epistemic norm. Impossible combinations will not be recommended as the best thing to do, regardless of how the individual states normatively perform by their own lights.

#### §7.4.3 Overdiscriminate state-differentiation

Finally, one of the main problems for all versions of doxastic normativism we have studied is that of overdiscriminating state-differentiation criteria. All formulations of belief norms that draw on intrinsic properties of belief entail that some mental states will count as beliefs and some won't, based on normative criteria. As we have seen, the proposals for what that normative criterion should be all lead to problematic cases of state exclusion: the judgments delivered by

normative definitions of belief do not track our intuitions about what counts as a belief. The problem of lawless deviance made this overdiscrimination effect particularly troublesome.

The knowledge-compatibility criterion in contrast is purely descriptive and draws on an uncontroversial feature of belief. The idea that belief involves a certain minimal threshold of confidence is well-known from the belief-credence-reduction literature and mirrored widely in our ordinary language practices ("I wouldn't say I believe it, but ..."). Furthermore, the fact that KC is not spelled out in terms of the abilities of the believing agent herself, but in terms of what would be possible for a hypothetical agent on the same affirmative strength, means that lawless deviance is unlikely to arise. There is no reason why an agent should be disqualified from believing for normative reasons. We considered even cases of creatures who might not be capable of knowledge at all and saw that they can nevertheless count as believers on knowledge-compatibility.

It is worthwhile to return to some of the examples of overdiscriminate normative criteria that we have encountered. For state normativism narrowly construed there is the problem that people with the wrong normative dispositions will be excluded from counting as believers. For conceptual normativism narrowly construed there is the exclusion of those who do not accept the proposed norm of belief. For teleologism there is the exclusion of people who have epistemically perverse aims. For teleo-functionalism there is the exclusion of mental states with the wrong evolutionary history.

All of these are symptoms of normativism's bridging strategy between the normative and the descriptive-psychological. Since the norm of belief is thought to originate from belief states or

from the concept of belief, we'd expect it to make a difference to actual belief formation.<sup>88</sup> This is why psychological data are often used to argue for normativism or for specific formulations of it (like the argument from the impossibility of believing at will, or the argument from transparency). The first challenge for the bridging strategy is to explain how norm guidance is supposed to work. We have discussed difficulties for specific proposals, like the strong norm motivation assumed by conceptual normativism (§4.2.3), or the over-intellectualisation of our mental dispositions risked by state normativism (§3.1.1). But furthermore, there is the question how it is possible that our beliefs are guided by the norm of belief in virtue of their nature while so many of them fail to be in accordance with epistemic norms.<sup>89</sup> The spirit of what Shah called the dilemma for teleologism generalises to normativism more widely: the dilemma between formulating the link between epistemic norms and the nature of belief strongly enough to capture our normative intuitions, but not so strong that it flies in the face of our empirical findings, or excludes belief-like phenomena from counting as beliefs.

On my account, there is a relationship between the epistemic norm and the psychology of belief as well: the evidential sensitivity of belief has to do with the fact that knowledge is a goal of inquiry, or an intrinsic value, or a source of epistemic obligation (the epistemic norm). Since subjects' supposed reaction to the epistemic norm is used to explain certain phenomena of belief formation, does my view not carry the same risk of normative state exclusion that the normative definitions of belief run?

---

<sup>88</sup> See the discussion of lawless deviance in chapter 5.

<sup>89</sup> And not just because the environment isn't epistemically conducive or misleading, but because of bias, haste, intellectual laziness etc.

#### §7.4.4 Normativity and psychology

I offered an account of the evidential sensitivity of belief that appealed to the idea that beliefs tend to occur in situation where we would judge ourselves to know. I argued that the reason for beliefs tending to occur in these circumstances is its knowledge-compatibility, along with the desirability of knowledge. Now the reader might think that that will open up the door for lawless deviance once more: will an agent who fails to value knowledge be disqualified from believing? No, because the desirability of knowledge, however we conceive of it normatively, is stipulated separately from agents' psychology or their concept of belief. An agent who does not subscribe to anything like the epistemic norm may well lack the concept of knowledge. But that does not change the fact that some of her attitudes are knowledge-compatible and others are not, since knowledge-compatibility is not agent-indexed.<sup>90</sup>

The acceptance or even grasping of the desirability of knowledge is not a condition on belief. It is however a factor that I argued makes a difference to how belief functions. So even though a subject's *capacity to believe* is not dependent on her relationship to the epistemic norm, the *way her beliefs behave* should be expected to be mediated by it. Is this a problem? Consider a subject who does not accept the epistemic norm, who is not at all inclined to think that knowledge is valuable or something she ought to strive for. Would we not indeed expect such a subject to be epistemically speaking worse at forming beliefs? Note that this kind of rejection of the value of knowledge would not amount simply to not having reflected on the matter or not being capable of forming epistemological judgments. The epistemic norm does not need

---

<sup>90</sup>Even if it were the case (I do not think it is) that failing to appreciate the epistemic norm would make her incapable of the state of knowledge (due to some condition on knowledge possession akin to Davidson's for belief).

to be accepted in a cognitively involved, conceptually drawn-out way. Someone simply valuing being non-luckily right in specific circumstances suffices for the norm to play the kind of role in belief formation I have in mind. To utterly reject the epistemic norm would amount to thinking one isn't better off knowing whether the drinking water is safe, whether one's spouse loves one, whether one's child is happy. A subject like that<sup>91</sup> might experience none of the unease of not taking herself to know the premises of her reasoning, and hence fail to match confidence to dispositional epistemic self-assessment, and hence form less evidentially sensitive beliefs. But this provides no reason to classify her as not capable of belief, and hence of falling outside of the scope of belief normativity: the only condition on belief is that it is knowledge-compatible, and her case provides no counterexample.

What about not caring about knowledge on the local level? Think about the kinds of cases David Papineau (2013) uses to argue against the existence of epistemic normativity: is it really valuable to have knowledge about the number of raindrops on my window pane or about whether my grandmother had toast for breakfast three days before her 23<sup>rd</sup> birthday? Two things can be said about these kinds of cases and their relevance to whether belief formation is mediated by the desirability of knowledge. Firstly, there is something odd to the idea of someone taking the necessary steps to form a belief about these two questions without caring about their truth at all. Since freedom to believe from neutral evidence is rare (and on involuntarism impossible), they would probably have to make at least a rough estimation of how many raindrops they can see, how large the window pane is etc, think about how often their grandmother tended to eat toast when they knew her, what time of year her birthday is,

---

<sup>91</sup> Where it's unclear we would grant such a subject as even possessing the concept of knowledge.

the availability of toast in the region where their 22-year-old grandmother lived etc. A subject who has considered the question and is even at all in a position to form a belief in response to it, is unlikely to genuinely not care about knowing. Secondly, it seems plausible that we are actually worse at forming reliable beliefs about topics we care very little about. So cases of belief that  $p$  without valuing the knowledge that  $p$  will be rare, and in the cases that do exist, we might even find evidence for the thesis that perceiving knowledge as desirable mediates evidential sensitivity of belief formation.

It is worth making explicit that the perceived desirability of knowledge that this explanation rests on does not depend on an intellectual capacity to reason about knowledge or to make normative judgments. It is very likely that non-human animals know things, and in so far as they know it seems clear that they value this. Many non-human animals can be observed showing unease about confusing or unexplored environments, about not previously tried kinds of food, about not previously encountered other animals. They do not only like to be safe, they like to know that they are safe. It is not only evidence of danger that causes them anxiety but lack of evidence for lack of danger as well. This behavioural and emotional pattern, of preferring situations in which one feels like one knows the premises one is entertaining or on which one is acting, over situations in which one does not, suffices for present purposes for counting as sensitive to the value of knowledge.

The knowledge-compatibility definition of belief does not make subjects' relationship to the epistemic norm a condition on belief possession. This relationship serves as an explanans of certain psychological phenomena, but these phenomena in turn are not conditions on belief possession. This approach makes for one significant advantage over normativism, since the

latter was shown to tend to exclude plausible belief-candidates from counting as beliefs. Another advantage is that the account of belief formation I have in mind rests on tendencies instead of on lawlike correlations: lowering confidence in cases of ignorance-self-ascription is something most agents tend to do, but it is not necessary that this always happens. Similarly, that one's actual mental states (first-order confidence) are in line with what they would be upon reflection (second-order epistemic judgment) is a tendency of the mind that can fail to manifest more or less often. Agents vary in their epistemological beliefs, implicit or explicit, about how much warrant they think is involved in knowledge. They also vary in how quickly they take themselves to know. The theory of belief presented in this chapter is better-suited than normativism to explain differences between epistemic agents in dimensions like need for closure, propensity to change one's mind, or effectiveness of epistemic self-censure.

## Conclusion

I set out to reject doxastic normativism, but in the macrostructure of this thesis I engaged mostly with four distinct versions of that view. Significant themes recurred throughout however. This conclusion will revisit some of them and reflect on the scope of the different objections that have been presented.

Some parts as of the thesis stayed close to texts with extant formulations of normativism. This was mostly the case for those parts that presented undercutting objections: Chapter 2 and the first half of Chapter 4. The undercutting work has been quite conclusive: three of the only arguments for doxastic normativism fail to establish that belief is normative. The argument from correctness rests on the false premiss that belief has truth as its only standard of correctness. Yet there is no evidence for that. The semantics of correctness ascription show the applicability of many standards of normative correctness to belief. Truth is one of them, but to assume that it is a privileged standard begs the question in an argument where belief's correctness condition is supposed to demonstrate what its norm is. The idea that truth is the only standard of correctness stems from a failure to disambiguate the word "correct": on the sense of the word that gives the correctness condition its ring of truism, the correctness condition is trivially true but also useless to the project of establishing anything normative about belief. The argument from correctness furthermore makes problematic assumptions about essences. This discussion also revealed a problem we would continue encountering: the challenge of state-differentiation. We saw that the (variable) correctness condition of belief is shared by the weak affirmative attitudes: they are said to be correct when true in the same circumstances as belief. Yet if it is the correctness condition that constitutes or grounds the norm of belief, that means that the

affirmative attitudes will be normatively indistinguishable. This is a problematic result, because we intuitively evaluate beliefs differently from weak affirmative attitudes. A norm of belief based on the correctness condition thus makes differences in epistemic normativity between distinct types of state difficult to account for.

The second argument for state normativism I addressed in Chapter 2 is the argument from dispositions. I discussed several avenues to resist this argument. Firstly, the premiss that essential dispositions have to be rational can be rejected if we adopt a view of rationality on which it is a function of the interaction of dispositions. Secondly, if we keep the notion of rationality employed in the premiss that rational dispositions are normative (because defeater-sensitive) fixed throughout the argument, the plausibility of the other premiss is seriously undermined. Thirdly, we saw that the argument assumes rationality to be binary, so that every disposition that is not rational, is irrational, which we do not have to accept. Later on in the thesis, I also mentioned worries for dispositionalism about mental states arising from the functionalist literature. In the discussion of the argument from dispositions, we encountered our first case of state exclusion: a hypothetical creature whose dispositions to form, maintain, and revise affirmative states without sensitivity to defeaters will not count as a believer because she will not count as rational. The binary rationality assumption that drives the argument from dispositions furthermore entails that this creature is entirely irrational, which is even more implausible than her incapacity for belief.

Looking back next to Chapter 4: the first half of this chapter again stayed close to a key text in the normativism literature: the argument from transparency for conceptual normativism narrowly construed. Though less substantially flawed than the arguments from correctness and

dispositions, the transparency argument too does not succeed in establishing normativism. Its failures stem firstly from the difficulty of construing the transparency principle in a way not subject to counterexamples, and secondly from its reliance on a strong norm motivation. The first problem mainly has to do with the formulation of the deliberative question that constitutes doxastic deliberation: if that question is "should I believe p or not believe p?", deliberation is not transparent to one's answer to the question "is p the case?", because one can suspend belief for other reasons. If on the other hand, the deliberative question is "should I believe p or believe not p?", transparency is no longer unique to belief, because deliberating about whether for instance to guess that p or to guess that not p is just as transparent to one's stance on p's truth. The second problem for the argument from transparency is due to the relationship between accepting a norm constitutive of a concept one is reasoning with and being internally motivated by that norm so strongly that one cannot deliberate in violation of it. That one can for instance deliberate about whether to make a promise one knows one is not going to keep shows that this strong norm motivation cannot simply be assumed, so the normativist has not done enough to show that her account of transparency works.

For rebutting objections, I did not always stay as close to specific formulations of normativist accounts. My two objections to state normativism, that it makes for difficulty with non-agential belief and that it leads to lawless deviance, were spelled out using dispositional state normativism, but could be brought to bear against it under different guises as well. Something I did not make explicit is that the non-agential belief objection is equally troubling for intentional teleologism as for state normativism. Conceptual normativism and teleo-functionalism on the other hand can account for non-agential belief easily (for conceptual normativism as long as the non-agents do not reason about belief, for teleo-functionalism as long as the non-agents

have a history that gave rise to their cognitive apparatuses). Lawless deviance was a problem we encountered for every version of normativism we studied. When I spelled the objection out in detail in Chapter 3, I mentioned reasons to think lawless deviance will be a consequence of doxastic normativism widely construed *generally*, not just for specific versions of it. The only exception would be a version of normativism that involves no norm effectiveness, so that the norm inherent to belief is not expected to make a difference to belief formation or to guide believers or cause phenomena of evidential sensitivity. That version would be free of the possibility of lawless deviance, but such a normativism is a less interesting view and is furthermore very difficult to argue for.

One of my rebutting objections to conceptual normativism, the possibility of conceptual disagreement, holds only against that formulation of normativism. The other, the state-differentiation challenge, appears to be systemic to normativism, although this stayed somewhat underdeveloped in the thesis, as I did not spell it out for every view we considered. Our theory of belief needs to differentiate belief from other affirmative attitudes. Doxastic normativism was meant to handle state-differentiation, but we saw that the opposite is the case. If it were the case that the norm of belief differentiates belief from other mental states (and some normativists have implied that it does), the problem of normative indistinguishability we have encountered in the dilemma of correctness conditions would translate into psycho-ontological indistinguishability as well. Chapter 4 looked at normative criteria other than correctness that could serve for state-differentiation, and found them unhelpful, because the difference in epistemic normativity between belief and weak affirmative attitudes was usually one of force rather than of criterion. The normative relationship between belief and truth and between belief and evidence exists for the weak affirmative attitudes as well, albeit with less

normative importance. Capturing this difference in degree of normative importance in a way usable for a definition of belief proved difficult.

The same state-differentiation problem could be spelled out in very similar terms against state normativism as well. For intentional teleologism, if belief were necessarily accompanied by a truth-aim (which I have argued it is not), there would be no reason why this could not also be the case for other affirmative attitudes like guessing. Indeed, Owens (2003) argues in detail that guessing is truth-aiming in the way Steglich-Petersen says belief is. It is also conceivable for there to be creatures that whenever they hope that  $p$ , do so accompanied by the intention to hope only if  $p$  is true. There is thus a similar state-differentiation challenge: the relationship between belief and truth is not unique. Contrary for instance to the relationship between shooting at sparrows and hitting sparrows with a bullet, where it seems that shooting at sparrows is the only activity that can be aim-directed to the goal of hitting sparrows with a bullet. Many things can be aimed at the truth. In light of that, explaining how belief differs from other affirmative (and even non-affirmative attitudes) through an aim is not self-evident.

For teleo-functionalism, it should also be clear that defining belief in a way that differentiates it from other attitudes is difficult. Just because belief has a given etiological function does not mean there could not be other states with the same function. Of course, the teleo-functionalism, who conceives of belief as a biological category, might not think it is the task of philosophers to uniquely define belief anyway (a sentiment I would be inclined to agree with). Nevertheless, an informative theory of belief should include a criterion for state-differentiation, and teleo-functionalism will not help provide it.

In the final chapter, I discussed another set of problems that goes to the heart of the normativist project at large (instead of to individual formulations of it): the challenge of obliging and state-overarching epistemic normativity. Obliging belief norms are firstly difficult to make sense of on a model of constitutive normativity. Furthermore, without a criterion of the belief norm that is unique to belief (which as we have seen is difficult to provide), obliging norms will entail massive overprescription. Foregoing obliging belief norms is similarly problematic however, because of underprescription. State-overarching epistemic norms finally, which recommend certain attitudes over others in specific situations, are impossible to provide on the normativist framework. Yet there is a robust intuition that that kind of normativity exists, which makes for poor prospects for the project of reducing the normativity of mental states to constitutive normativity.

We have thus seen some problems for specific formulations of and arguments for doxastic normativism, and we have seen others that are general problems systemic to the view at large. Much of the final chapter offered constructive work, building on the failures of normativism. The problem of normative indistinguishability means we need to find a source for the norm of belief that is truly unique to belief. Lawless deviance means we need to find a criterion of state-differentiation that does not lead to state exclusion due to normative aberrance. The knowledge-compatibility (KC) principle is an answer to both of these challenges. Firstly, only things that are beliefs are knowledge-compatible, which explains belief's unique normative status. KC is also compatible with the weaker normative status of the weak affirmative attitudes. And secondly, knowledge-compatibility is a non-normative property, meaning there is no amount of epistemic deviance that will lead to state exclusion.

I have treated doxastic normativism as a certain strategy of defining belief: one on which norms of belief follow from its essence, either because they are part of that essence or through entailment. I talked throughout the thesis of normativist “definitions” of belief, but I mentioned at the outset that normativism is compatible with there not being a definition of belief. And many extant accounts of normativism do not attempt a full definition. What they offer however, are principles about belief that would have to be part of such a definition were we to formulate one. It is this essential claim that is at the heart of normativism: belief would not be what it is if it were not governed by the norm of belief. But then in the account of belief that I formulated to solve problems identified for normativism, I did not put forward a constitutive essence of belief myself<sup>92</sup>. Is this a slight of hand? Is perhaps the real problem with normativism that it seeks out essences of belief at all, not what its candidate essence is?

First of all, I called only such authors “normativist” who explicitly speak of essential or constitutive norms of belief. While I think the ultimate reason for the failure of normativism is that it stipulates a *normative* essence, not that it stipulates an essence *tout court*, regardless of what this ultimate reason behind its difficulties is, my arguments against it stand. Furthermore though, even if we were to reformulate some normativist proposals to make only claims about normative principles that are necessary or sufficient for belief (without mentioning essences), difficulties would persist. My rebuttal of the correctness condition is a direct case in point: it is not just false that the correctness condition is essential to belief; the correctness condition is false itself. Turning to state-differentiation, there is not just a lack of a normative property *essential* to belief that

---

<sup>92</sup> The knowledge-compatibility might well be essential to the mental state, but I am not claiming that it is in this thesis. It is unlikely that it would be essential to the concept of belief.

can distinguish it from weak affirmative attitudes, there is a lack of *any* normative property that suffices for that task.<sup>93</sup> Some of the problems of epistemic normativity that I raised in Chapter 7 are problems for non-normativist belief norms as well. For instance, on a truth or evidence criterion, the norm of belief will either over- or underprescribe. This is not due to constitutivism but to the state-centric approach to epistemic normativity<sup>94</sup>. The problem of state-overarching normativity is similarly ultimately due to state-centrism. All of this warrants the conclusion that it is not merely its insistence on *essential* normative properties that compromises doxastic normativism, but its commitment to normative properties of belief generally. The advantage the knowledge-first account has over normativism is not simply due to its not setting itself the difficult task of unearthing essences. Its main advantage is that it puts no necessary normative conditions on belief and provides a non-normative sufficient condition.

The knowledge-first account sought to demarcate a property unique to states that are beliefs and to explain some of the psychological phenomena that normativists are interested in as well as the epistemic normativity of affirmative attitudes. The effectiveness of knowledge-compatibility as a state-differentiating criterion with psychological and normative explanatory power is one more reason against doxastic normativism. If a descriptive account can meet the theoretical goals that were motivations for a normative account (that failed to meet those goals), that weighs against the latter. One specific advantage of a descriptive account is its compatibility with a much broader range of scenarios about the nature of the mind, specifically with respect

---

<sup>93</sup> This does not follow from the discussion in Chapter 4 by itself, but draws on state-overarching epistemic normativity as well.

<sup>94</sup> State-centrism and normativism complement one another, but views that are not normativist can still be state-centrist.

to its tendency towards rationality. The psychology of irrationality began being seriously developed in the 1970s, through for instance cognitive bias research<sup>95</sup>. The list of cognitive biases subject to research is now very long. Additionally, the notion is gaining ground that many of these biases might in fact be adaptations, which do not indicate that the mind is irrational, but instead that traditional rationality is maladaptive.<sup>96</sup> But of course the epistemic norms researched by the philosophers we encountered are norms of that “traditional” rationality. So however evolutionary psychologists may evaluate the abundance of biases in the human mind, that abundance bears negatively on the likelihood of the human mind being largely rationality-conforming. The rules of rationality are rules to guide believers by recommending or censuring beliefs. They are not *modi operandi* of belief formation. Indeed, we have encountered throughout this thesis examples of beliefs that were produced by *modi operandi* very far removed from the rules of rationality. In order to account for the range of rationality and irrationality that the real subjects of psychology and the hypothetical believers of thought experiments display, it is helpful to have an account of epistemic normativity that is rooted not in what belief is like, but in what we need it for. The last chapter included a first step in that direction. On the picture I sketched, it matters what we believe not because of the nature of belief itself. It matters because knowledge matters. And belief is our only shot at knowledge.

---

<sup>95</sup>Kahneman and Tversky wrote several texts that proved foundational of such research programmes (Kahneman & Tversky, 1972, 1983).

<sup>96</sup> See Haselton et al. (2009) for a review.



## Bibliography

- Anscombe, G. E. M. (1957). *Intention*. Harvard University Press.
- Audi, R. (2001). Doxastic Voluntarism and the Ethics of Belief. In *Knowledge, Truth, and Duty* (pp. 93–108). Oxford University Press.
- Basolo, A. L. (1995). Phylogenetic Evidence for the Role of a Pre-Existing Bias in Sexual Selection. *Proceedings of the Royal Society B: Biological Sciences*, 259(1356), 307–311.
- Basolo, A. L., & Alcaraz, G. (2003). The turn of the sword: length increases male swimming costs in swordtails. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1524), 1631–1636.
- Basolo, A. L., & Wagner, W. E. (2004). Covariation between predation risk, body size and fin elaboration in the green swordtail, *Xiphophorus helleri*. *Biological Journal of the Linnean Society*, 83(1), 87–100.
- Bayne, T., & Pacherie, E. (2005). In defence of the doxastic conception of delusions. *Mind & Language*, 20(2), 163–188.
- Bergmann, M. (2006). *Justification without Awareness*. Oxford University Press.
- Bird, A. (2007). Justified Judging. *Philosophy and Phenomenological Research*, 74(1), 81–110.
- Bisazza, A. (1993). Male competition, female mate choice and sexual size dimorphism in poeciliid fishes. *Marine Behaviour and Physiology*, 23(1–4), 257–286.
- Boghossian, P. A. (2003). The Normativity of Content. *Philosophical Issues*, 13(1), 31–45.
- Boghossian, P. A. (2008). Epistemic Rules. *Journal of Philosophy*, 105(9), 472–500.
- Boyce, K., & Moon, A. (2016). In defense of proper functionalism: cognitive science takes on Swampman. *Synthese*, 193(9), 2987–3001.
- Braithwaite, R.B. (1932–1933). The nature of believing. *Proceedings of the Aristotelian Society*, 33: 129–146.
- Brandt, R. (2001). Modality, Normativity, and Intentionality. *Philosophy and Phenomenological Research*, 63(3), 587–609.
- Bykvist, K., & Hattiangadi, A. (2007). Does Thought Imply Ought? *Analysis*, 64(4).
- Bykvist, K., & Hattiangadi, A. (2013). Belief, Truth, and Blindspots. in Chan, Timothy (ed). *The Aim of Belief*. Oxford: Oxford University Press, p. 100–122
- Camp, E. (2009). A language of baboon thought? In R. Lurz (Ed.), *Philosophy of Animal Minds*. New York: Cambridge University Press.
- Carrier, David, Kapoor, T. K., Kimura, Tasuku, Nickels, Martin K., Satwanti, Scott, Eugenie C., So, Joseph K., & Trinkaus, Erik (1984). The Energetic Paradox of Human Running and Hominid Evolution [and Comments and Reply], *Current Anthropology* 25 (4), 483–495

- Carruthers, P. (2004). On being simple minded. *American Philosophical Quarterly*, 41, 205–220.
- Clark, C. J., & Dudley, R. (2009). Flight costs of long, sexually selected tails in hummingbirds. *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1664), 2109–2115.
- Copp, D. (2009). Towards a Pluralist and Teleological Theory of Normativity. *Philosophical Issues*, 19(1), 21–37.
- Currie, G. (2000). Imagination, delusion and hallucinations. *Mind & language*, 15(1), 168–183.
- Davidson, D. (1975). Thought and talk. In S. D. Guttenplan (Ed.), *Mind and Language*. Clarendon Press.
- Davidson, D. (1982). Rational Animals. *Dialectica*, 36, 317–328.
- Davidson, D. (1987). Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association*, 66(1), 441–458.
- Dennett, D. (1981). Intentional Systems Theory. In *Brainstorms* (p. 377).
- Dennett, D. (1995). Do Animals Have Beliefs? In H. Roitblat & J. Meyer (Eds.), *Comparative Approaches to Cognitive Science*. MIT Press.
- Dennett, D. (2009). Intentional Systems Theory. In A. Beckermann & S. Walter (Eds.), *The Oxford Handbook of Philosophy of Mind* (pp. 1–14). Oxford University Press.
- Douven, I. (2009). Assertion, Moore, and Bayes. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 144, 361–375.
- Engel, P. (2000). Wherein Lies the Normative Dimension in Meaning and Mental Content? *Philosophical Studies*, 100, 305–321.
- Engel, P. (2001). Is truth a norm? In P. Kotatko, P. Pagin, & G. Segal (Eds.), *Interpreting Davidson* (pp. 37–51). University Of Chicago Press.
- Engel, P. (2002). Free believers. *Manuscrito*, 25(3), 155–175.
- Engel, P. (2004). Truth and the Aim of Belief. *Laws and Models in Science*, 77–97.
- Engel, P. (2009). Epistemic responsibility without epistemic agency. *Philosophical Explorations*, 12(2), 205–219.
- Engel, P. (2013a). Belief and the right kind of reason. *Teorema*, 32(3), 19–34.
- Engel, P. (2013b). I-Doxastic Correctness. *Aristotelian Society Supplementary Volume*, 87(1), 199–216.
- Engel, P. (2013c). In Defence of Normativism about the Aim of Belief. In T. Chan (Ed.), *The Aim Of Belief*.
- Fassio, D. (forthcoming.). Belief, Correctness and Constitutivity. *European Journal of Philosophy*.
- Fassio, D. (2011). Belief, Correctness, and Normativity. *Logique et Analyse*, 54(216), 471–486.
- Fassio, D. (2016). A Problem for Deontic Doxastic Constitutivism. *Philosophical Papers*, 34(3), 343–364.

- Fine, K. (1994). Essence and Modality: The Second Philosophical Perspectives Lecture. *Philosophical Perspectives*, 8, 1.
- Fine, K. (1995). Senses of Essence. In W. Sinnott-Armstrong, D. Raffman, & N. Asher (Eds.), *Modality, morality, and belief: essays in honor of Ruth Barcan Marcus* (p. 270).
- Frost, K. (2014). On the Very Idea of Direction of Fit. *Philosophical Review*, 123(4), 429–484.
- Gendler, T. S. (2008). Alief in action (and reaction). *Mind & Language*, 23(5), 552–585.
- Giardina, B., Messina, I., Scatena, R., & Castagnola, M. (1995). The Multiple Functions of Hemoglobin. *Critical Reviews in Biochemistry and Molecular Biology*, 30(3), 165–196.
- Gibbard, A. (1994). Meaning and Normativity. *Philosophical Issues*, 5, 95–115.
- Gibbard, A. (2003). Thoughts and Norms. *Noûs-Supplement: Philosophical Issues*, 13, 83–98.
- Gibbard, A. (2005). Truth and Correct Belief. *Philosophical Issues*, 15(1), 338–350.
- Glock, H. J. (2010). Can animals judge? *Dialectica*, 64, 11–33.
- Glüer, K., & Wikforss, Å. (2009). Against Content Normativity. *Mind*, 118(469), 31–70.
- Glüer, K., & Wikforss, Å. (2010). The truth norm and guidance: A reply to Steglich-Petersen. *Mind*, 119(475), 757–761.
- Glüer, K., & Wikforss, Å. (2013a). Against Belief Normativity. In T. Chan *The Aim Of Belief* (pp. 121–146). Oxford University Press.
- Glüer, K., & Wikforss, Å. (2013b). Aiming at truth: On the role of belief. *Teorema*, 32(3), 137–162.
- Glüer, K., & Wikforss, Å. (2015). Meaning Normativism: Against the Simple Arguer. *Organon F* 22 (Supplementary Issue) 63–73
- Graham, P. J. (2014a). Functions, epistemic warrant and natural norms. In A. Fairweather & O. Flenagan (Eds.), *Naturalizing Epistemic Virtue*. Cambridge University Press.
- Graham, P. J. (2014b). Warrant, Functions, History. In A. Fairweather & O. Flenagan (Eds.), *Naturalizing Epistemic Virtue*. Cambridge University Press.
- Gregory, A. (2012). Changing Direction on Direction of Fit. *Ethical Theory and Moral Practice*, 15(5), 603–614.
- Haselton, M. G., Bryant, G. A., Wilke, A., Frederick, D. A., Galperin, A., Frankenhuis, W. E., & Moore, T. (2009). Adaptive Rationality: An Evolutionary Perspective on Cognitive Bias. *Social Cognition*, 27(5), 733–763.
- Hattiangadi, A. (2006). Is Meaning Normative? *Mind and Language*, 21, 220–240.
- Hawthorne, J., Rothschild, D., & Spectre, L. (2016). Belief is weak. *Philosophical Studies*, 173(5), 1393–1404.
- Humberstone, I. L. (1992). Direction of fit. *Mind*, 101(401), 59–83.
- Kahneman, D., & Tversky, A. (1972). Subjective Probability: A Judgment of Representativeness.

*Cognitive Psychology*, 3(430–454).

- Kahneman, D., & Tversky, A. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293–315.
- Kelp, C. Assertion (forthcoming): A Function First Account. *Noûs*.
- Kripke, S. (1982). *Wittgenstein on Rules and Private Language*. Harvard University Press.
- Kruglanski, A., Webster, D. M., & Klem, A. (1993) Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of Personality and Social Psychology*, 65(5), 861–876.
- Littlejohn, C. (2013). The Russellian Retreat. *Proceedings of the Aristotelian Society*, 113(3), 293–320.
- Low, P. (2012). *The Cambridge Declaration on Consciousness*. Cambridge.
- Lynch, M. (2009). The Value of Truth and the Truth of Values. In A. Haddock, A. Millar, & D. Pritchard (Eds.), *Epistemic Value*. Oxford University Press.
- Marcus, R. B. (1991). Some revisionary proposals about belief and believing. In *Causality, method, and modality* (pp. 143–173). Springer, Dordrecht.
- Martin, E., & Hine, R. (2008). *A Dictionary of Biology*. Oxford University Press.
- McHugh, C. (2012a). Belief and Aims. *Philosophical Studies*, 160(3), 425–439.
- McHugh, C. (2012b). The truth norm of belief. *Pacific Philosophical Quarterly*, 93(1), 8–30.
- McHugh, C. (2014). VIII - Fitting belief. *Proceedings of the Aristotelean Society*, 114(2), 167–187.
- McHugh, C., & Whiting, D. (2014). The Normativity of Belief. *Analysis*, 74(4), 698–713.
- McLaughlin, B. (2000). *What Functions Explain*. Cambridge University Press.
- McLaughlin, B. P., & Cohen, J. D. (Eds.). (2007). *Contemporary Debates in Philosophy of Mind*. Blackwell.
- Miller, G. (2000). Sexual selection for indicators of intelligence. *Novartis Foundation Symposium*, 233, 260–70–80.
- Millikan, R. G. (1984). *Language, Thought, And Other Biological Categories*. MIT Press.
- Millikan, R. G. (1986). Thoughts Without Laws; Cognitive Science with Content. *The Philosophical Review*, 96(1), 47–80.
- Millikan, R. G. (1991). Speaking Up for Darwin. In B. Loewer & G. Rey (Eds.), *Meaning in Mind: Fodor and his critics* (pp. 151–165). Cambridge, MA: Blackwell.
- Millikan, R. G. (1993). *White Queen Psychology and Other Essays for Alice*. MIT Press.
- Miyazono, K., & Bortolotti, L. (2014). The causal role argument against doxasticism about delusions. *Avant: Trends in Interdisciplinary Studies*, 30–50.
- Neander, K. (1995). Misrepresenting & Malfunctioning. *Philosophical Studies*, 79(2), 109–141.

- Neander, K. (1996). Swampman Meets Swampcow. *Mind and Language*, 11(1), 70–130.
- Nolfi, K. (2015). How to Be a Normativist About the Nature of Belief. *Pacific Philosophical Quarterly*, 96(2015), 181–204.
- Nottelmann, N. (2007). *Blameworthy belief: a study in epistemic deontologism*. Springer.
- Ostrom, T. M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology*, 5(1), 12–30.
- Owens, D. J. (2003). Does belief have an aim? *Philosophical Studies*, 115(3), 283–305.
- Papineau, D. (1999). Normativity and judgment. *Proceedings of the Aristotelean Society*, 73(16–43).
- Papineau, D. (2013). There Are No Norms Of Belief. In T. Chan (Ed.), *The Aim Of Belief*.
- Plantinga, A. (1993). *Warrant and Proper Function*. New York: Oxford University Press.
- Pojman, L. P. (1999). Believing, Willing, and the Ethics of Belief. In L. P. Pojman (Ed.), *The Theory of Knowledge*. Wadsworth.
- Price, C. (1998). Determinate Functions. *Nous*, 32(1), 54–75.
- Radford, C. (1966). Knowledge - By Examples. *Analysis*, 27(1), 1–11.
- Raleigh, T. (2013). Belief norms and blindspots. *Southern Journal of Philosophy*, 51(2), 243–269.
- Rescorla, M. (2009). Chrysippus' dog as a case study in non-linguistic cognition. In R. Lurz (Ed.), *The Philosophy of Animal Minds*. New York: Cambridge University Press.
- Rose, D., Buckwalter, W., & Turri, J. (2014). When Words Speak Louder than Actions: Delusions, Belief, and the Power of Assertion. *Australasian Journal of Philosophy*, 92(4), 683–700.
- Rose, D., & Schaffer, J. (2013). Knowledge Entails Dispositional Belief. *Philosophical Studies*, 166(supplement 1), 19–50.
- Rosen, G. (2015). Real Definition. *Analytic Philosophy*, 56(3), 189–209.
- Rott, H. (2017). Negative Doxastic Voluntarism and the concept of belief. *Synthese*, 194(8), 2695–2720.
- Ryle, G. (1949). *The concept of mind*. New York: Barnes & Noble
- Schroeder, T. (2003). Donald Davidson's Theory of Mind is Non-Normative. *Philosopher's Imprint*, 3 (1), 1-14.
- Schroeder, M. (2015). Is Knowledge Normative? *Philosophical Issues*, 25(1), 379–395.
- Schwitzgebel, E. (2002). A phenomenal, dispositional account of belief. *Nous*, 36(2), 249-275.
- Schwitzgebel, E. (2010). Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly*, 91(4), 531-553.
- Schwitzgebel, E. (2012). Mad belief?. *Neuroethics*, 5(1), 13-17.
- Scott-Kakures, D. (1994). On Belief and Captivity of the Will. *Philosophy and Phenomenological*

*Research*, 54, 77–103.

- Shah, N. (2003). How Truth Governs Belief. *Philosophical Review*, 112(4), 447–482.
- Shah, N. (2013). Why we reason the way we do. *Philosophical Issues*, 23, 311–325.
- Shah, N., & Velleman, D. (2005). Doxastic deliberation. *Philosophical Review*, 114(4), 497–534.
- Simion, M. (2015). Perception, History and Benefit. *Episteme*, 13(1), 61–76.
- Smith, M. (1987). The Humean theory of motivation. *Mind*, 96(381), 36–61.
- Smithies, D. (2012a). Moore's Paradox and the Accessibility of Justification. *Philosophy and Phenomenological Research*, 85, 273–300.
- Smithies, D. (2012b). The normative role of knowledge. *Noûs*, 46(2), 265–288.
- Steglich-Petersen, A. (2006). No norm needed: On the aim of belief. *Philosophical Quarterly*, 56, 499–516.
- Steglich-Petersen, A. (2010). The Truth Norm and Guidance: a Reply to Gluer and Wikforss. *Mind*, 119(475), 749–755.
- Stich, S. (1978). Do Animals Have Beliefs? *Australasian Journal of Philosophy*, 57, 15–28.
- Sutou, S. (2012). Hairless mutation: a driving force of humanization from a human–ape common ancestor by enforcing upright walking while holding a baby with both hands. *Genes to Cells*, 17, 264–272
- Toppinen, T. (2015). How Norms (Might) Guide Belief. *Philosophical Studies*, 23(3), 396–409.
- Tumulty, M. (2011). Delusions and dispositionalism about belief. *Mind & Language*, 26(5), 596–628.
- Vahid, H. (2006). Aiming at Truth: Doxastic vs. Epistemic Goals. *Philosophical Studies*, 131, 303–335
- Van Hiel, A., & Mervielde, I. (2003). The Need for Closure and the Spontaneous use of Complex and Simple Cognitive Structures. *The Journal of Social Psychology*, 143(5), 559–568.
- Velleman, D. (2000a). On the Aim of Belief. In *The Possibility of Practical Reasoning*. Oxford University Press.
- Velleman, D. (2000b). *The Possibility of Practical Reason*. Oxford University Press.
- Webster, D. M., & Kruglanski, A. (1994). Individual Differences in Need for Closure. *Journal of Personality and Social Psychology*, 67, 1049–1062.
- Wedgwood, R. (2002). The Aim of Belief. *Philosophical Perspectives*, 16, 267–297.
- Wedgwood, R. (2007a). Normativism Defended. In B. McLaughlin & J. D. Cohen (Eds.), *Contemporary Debates in Philosophy of Mind* (pp. 85–102). Blackwell.
- Wedgwood, R. (2007b). *The Nature of Normativity*. New York: Oxford University Press.
- Wedgwood, R. (2009). The Normativity of the Intentional. In A. Beckerman, B. McLaughlin, & S. Walter (Eds.), *The Oxford Handbook of the Philosophy of Mind* (pp. 421–437). Oxford

University Press.

- Whiting, D. (2007). The Normativity of Meaning Defended. *Analysis*, 67, 133–140.
- Whiting, D. (2010). Should I Believe the Truth. *Dialectica*, 64(2), 213–224.
- Whiting, D. (2012) Does Belief Aim (Only) At The Truth? *Pacific Philosophical Quarterly*, 93, 279-300
- Whiting, D. (2013). Truth: The Aim and Norm of Belief. *Teorema*, 32(3), 121–135.
- Wilkinson, S. (2013). The status of delusion in the light of Marcus's 'Revisionary proposals'. *Theoria*, 28 (3), pp. 421-436.
- Williams, B. (1970). Deciding to Believe. In H. E. Kiefer & M. K. Munitz (Eds.), *Language, Belief, and Metaphysics* (pp. 95–111). Albany: SUNY Press.
- Williams, B. (1973). Deciding to Believe. In *Problems of the Self*. Cambridge University Press.
- Williams, J. N. (2007). The Surprise Exam Paradox: Disentangling Two Reductios. *Journal of Philosophical Research*, 32, 67–95.
- Williamson, T. (2002). *Knowledge and its Limits*. Oxford University Press.
- Williamson, T. (2006). Conceptual truth. In *Aristotelian Society Supplementary Volume* (Vol. 80, No. 1, pp. 1-41).
- Williamson, T. (2008). *The philosophy of philosophy*. John Wiley & Sons.
- Wright, L. (1973). Functions. *Philosophical Review*, 82(2), 139–168.
- Zangwill, N. (1998). Direction of Fit and Normative Functionalism. *Philosophical Studies*, 91(2), 173–203.
- Zangwill, N. (2005). The Normativity of the Mental. *Philosophical Explorations*, 8(1), 1–19.
- Zangwill, N. (2010). Normativity and metaphysics of mind. *Australasian Journal of Philosophy*, 88(1), 21–39.
- Zimmerman, A. (2007). The nature of belief. *Journal of Consciousness Studies*, 14(11), 61-82.



# Index

|   |     |               |        |
|---|-----|---------------|--------|
| Def. 1.1 affirmative attitudes.....                     | 19  | Ex. 1.1 ..... | 19     |
| Def. 1.10 teleo-functionalism .....                     | 33  | Ex. 2.1 ..... | 57     |
| Def. 1.11 teleo-functional norm of belief.              | 33  | Ex. 2.2 ..... | 57     |
| Def. 1.2 doxastic normativism .....                     | 26  | Ex. 2.3 ..... | 57     |
| Def. 1.3 essential doxastic normativism....             | 27  | Ex. 2.4 ..... | 60     |
| Def. 1.4 conceptual doxastic normativism .....          | 28  | Ex. 2.5 ..... | 61     |
| Def. 1.5 doxastic normativism (narrowly construed)..... | 29  | Ex. 2.6 ..... | 61, 63 |
| Def. 1.6 doxastic normativism (widely construed) .....  | 29  | Ex. 2.7 ..... | 68     |
| Def. 1.8 teleologism .....                              | 32  | Ex. 3.1 ..... | 78     |
| Def. 1.9 teleological norm of belief .....              | 33  | Ex. 3.2 ..... | 78     |
| Def. 4.1 Strong Belief-Action.....                      | 146 | Ex. 3.3 ..... | 79     |
| Def. 7.1 knowledge-compatibility .....                  | 205 | Ex. 3.4 ..... | 97     |
| Def. 7.2 very strong freedom to believe                 | 214 | Ex. 3.5 ..... | 98     |
| Def. 7.3 strong freedom to believe.....                 | 214 | Ex. 3.6 ..... | 99     |
| Def. 7.4 strong freedom from belief.....                | 214 | Ex. 3.7 ..... | 100    |
| Def. 7.5 weak freedom to believe .....                  | 214 | Ex. 3.8 ..... | 101    |
| Def. 7.6 weak freedom from belief.....                  | 214 | Ex. 3.9 ..... | 104    |
| Def. 7.7 very weak freedom to believe..                 | 215 | Ex. 4.1 ..... | 121    |
| Def. 7.8 very weak freedom from belief:                 | 215 | Ex. 4.2 ..... | 121    |
|   |     | Ex. 4.3 ..... | 131    |
|   |     | Ex. 4.4 ..... | 148    |
|   |     | Ex. 4.5 ..... | 148    |
|   |     | Ex. 4.6 ..... | 149    |
|   |     | Ex. 5.1 ..... | 163    |
|   |     | Ex. 5.2 ..... | 172    |
|   |     | Ex. 6.1 ..... | 196    |
|   |     | Ex. 6.2 ..... | 199    |
|   |     | Ex. 7.1 ..... | 206    |
|   |     | Ex. 7.15..... | 207    |
|   |     | Ex. 7.16..... | 207    |
|   |     | Ex. 7.17..... | 209    |
|   |     | Ex. 7.18..... | 209    |
|   |     | Ex. 7.2 ..... | 206    |
|   |     | Ex. 7.3 ..... | 206    |
|   |     | Ex. 7.4 ..... | 206    |
|   |     | Ex. 7.5 ..... | 206    |
|   |     | Ex. 7.6 ..... | 207    |
|   |     | Ex. 7.7 ..... | 207    |

acceptance (of a norm)..... 89, 125  
action ..... 145, 150, 152, 227  
affirmative attitudes19, 20, 60, 123, 133,  
149, 153, 203, 215, 223, 229  
    semantics of..... 209  
agnosticism68, 120, 122, 138, 155, 213, 224,  
231  
aim of belief ..... 159, 161, 164, 172, 230  
    by extension.....166, 169, 173  
    intentional..... *See* intention  
    non-alethic ..... 173  
    normative ..... 159  
    sub-intentional ..... 161  
aim-directed activities114, 159, 162, 165,  
167, 168  
assertion ..... 106  
assumption .....*See* affirmative attitudes  
belief  
    animal ..... *See* non-agential  
    concept of111, 116, 124, 127, 140, 142,  
    174, 210, 230, 237  
    dispositional..... 85, 151  
    evolution of..... 186  
    infant..... *See* non-agential  
    non-agential ..... 75, 81, 83, 86, 151  
    semantics of.....58  
biological categories..... 88, 187  
brains in vats ..... 152  
censure .37, 41, 79, 103, 132, 170, 201, 230  
commitment.....21, 203, 205  
confidence ..... 21, 203, 205, 221, 235, 240  
control (over mental states)123, 213, 215,  
224  
correctness  
    dilemma of..... 62, 133  
    of affirmative attitudes .....61  
    semantics of.....54  
correctness condition .44, 52, 124, 133, 135  
credences.....22, 203, 208  
defeater-sensitivity .....65, 68, 76  
definition .....31  
deviant states ..... 104, 110, 127, 175, 236  
    aberrant history..... 200  
    bias ..... 104, 221  
    careless .....131, 239  
    failure to value knowledge.....237  
    hallucination .....101  
    phobia..... 119, 121, 221  
    refusal to believe.....231  
    self-deception.....132, 172  
    widespread deviance .....174  
    wishful thinking100, 116, 118, 162, 166,  
    193  
direction of fit..... 17  
disagreement .....128, 131  
dispositionalism about belief ..... 65, 81, 87  
dispositions ..... 68, 91  
    essential..... 65  
    rational..... 65, 66, 69, 87, 96, 102  
    to act.....145, 150, 152, 154  
    to believe .....222  
doxastic deliberation112, 116, 119, 120,  
127, 130  
doxastic involuntarism40, 212, 215, 223,  
225, 236  
doxastic voluntarism*See* doxastic  
    involuntarism  
epistemic normativity.....39, 226  
    agent-centric..... 218, 224, 227, 232  
    general epistemic norm218, 224, 228,  
    239  
    of belief.....226  
    of weak affirmative attitudes107, 138,  
    141, 153, 212, 226, 227, 231  
    primary norm..... 77  
    state-centric ..... 218, 224, 229, 234  
    state-overarching .....142, 233  
essence ..... 29, 47, 211  
    consequential..... 30  
    constitutive ..... 30  
estimation..... *See* affirmative attitudes  
evaluative ..... *See* axiological  
evidential sensitivity114, 117, 126, 143, 211,  
216, 220, 222, 236, 238  
    of weak affirmative attitudes ..... 143  
evidentialism... *See* norm of belief:evidential  
function  
    conflict .....190  
    dispositional ..... 180

etiological ..... 182  
 indeterminacy ..... 188  
 malfunctioning ..... 88  
 normative ..... 88, 179, 181, 192  
 of belief ..... 177, 182  
 plurality ..... 189, 192  
 shifting ..... 195, 196  
 typing ..... 181, 188, 190, 191, 197, 201  
 functional definition .... *See* function (typing)  
 games ..... 93, 106, 230  
 guessing ..... *See* affirmative attitudes  
 guidance (by a norm) 89, 91, 92, 95, 96, 97,  
 102, 103, 236  
 hypothesis ..... *See* affirmative attitudes  
 individuation ..... *See* state differentiation  
 intention ..... 160, 162, 166, 168  
 intentionality ..... 16, 28, 64, 107, 192, 198  
 introspection ..... 150, 228  
 investigation ..... 153, 219, 227  
 irrational ..... *See* rationality:irrational  
 knowledge ..... 217  
     desirability of ..... 218, 225, 237  
     normative significance of ..... 217  
 knowledge-compatibility ..... 207, 232, 235  
 lawless deviance 95, 103, 105, 107, 130, 169,  
 172, 201, 235, 237  
 minimal discrepancy ..... 221, 240  
 Moore paradox ..... 221  
 motivation (by a norm) ..... 126, 160  
 naturalism ..... 160  
 need-for-closure ..... 144  
 norm of belief  
     authority ..... 25, 103, 212  
     criterion 23, 129, 138, 139, 141, 204, 212,  
     231  
     evidential ..... 139, 144, 204  
     logical form ..... 24, 230  
     source ..... 24, 103, 232  
     species ..... 23  
 norms  
     axiological ..... 24, 50, 177, 192  
     categorical ..... 41  
     deontic ..... 24  
     effectiveness ..... 93, 162, 236  
     epistemic ..... *See* epistemic normativity  
     force ..... 139, 141, 165, 212  
     functional ..... 24, 33, 41, 178  
     instrumental 24, 33, 41, 160, 169, 170, 171  
     intrinsic ..... 34  
     obliging ..... 50, 138, 229  
     permissive ..... 50, 138, 230  
     thickness ..... 37, 81, 178  
 ought implies can ..... 40, 77  
 overdiscriminate state differentiation .... *See*  
     state exclusion  
 overprescription ..... 138, 141, 231  
 practical interests ..... *See* stakes  
 practical reasoning ..... *See* action  
 rationality ..... 65, 67, 70, 76  
     binary ..... 70, 72  
     irrationality ..... 38, 70, 71, 93, 102, 110  
 scepticism  
     normative ..... 25, 103  
     Pyrrhonian ..... 139  
 stakes ..... 150  
 state differentiation 47, 62, 107, 112, 133,  
 135, 201, 211  
     non-normative ..... 143  
 state exclusion 71, 80, 87, 102, 130, 164,  
 198, 201, 210, 234  
 state indistinguishability  
     normative ..... 133, 136  
     psycho-ontological ..... 134, 135  
 stereotype ..... 174  
 suspension of belief ..... *See* agnosticism  
 suspicion ..... *See* affirmative attitude  
 Swampman ..... 198  
 teleosemantics ..... 27, 177, 189, 198  
 trait ..... 183, 184, 192, 195, 197  
     selected ..... 185  
     typing ..... 187  
 transparency 112, 115, 118, 120, 123, 133,  
 136, 212, 236  
     dilemma of ..... 116  
     of affirmative attitudes ..... 123  
 truth regulation ..... *See* evidential sensitivity  
 underprescription ..... 138, 141, 230