

Dissertation

The Problem of Mental Responsibility.

Outlines of an Ethics of Mind

Der philosophischen Fakultät mit Fachbereich Theologie

Dem Fachbereich Philosophie

zur

Erlangung des Doktorgrades Doctor philosophiae (Dr. phil.)

vorgelegt von

Sebastian Schmidt, M.A.

Als Dissertation genehmigt

von der philosophischen Fakultät mit Fachbereich Theologie

und vom Fachbereich Philosophie

der Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung: 13.11.2020

Vorsitzender des Promotionsorgans: Prof. Dr. Thomas Demmelhuber

Gutachter:

Prof. Dr. Gerhard Ernst (FAU Erlangen-Nürnberg)

Prof. Dr. Christian Seidel (KIT Karlsruhe)

Prof. Dr. Anne Meylan (Universität Zürich)

Table of Contents

TABLE OF CONTENTS	III
<u>0 INTRODUCTION</u>	<u>VII</u>
0.1 THE PROJECT	IX
0.2 SOME INITIAL CONCEPTUAL CLARIFICATIONS	XIII
0.2.1 NORMS, STANDARDS, RULES, AND REQUIREMENTS	XIV
0.2.2 NORMS AND REASONS	XVI
0.2.3 RESPONSIBILITY AND BLAME	XX
0.2.4 RATIONALITY AND REASONS	XXIII
0.3 SUMMARY OF CHAPTERS	XXVIII
<u>PART I THE PROJECT AND THE PROBLEM</u>	<u>1</u>
<u>1 SOME HISTORICAL BACKGROUND</u>	<u>1</u>
1.1 THE ETHICS OF BELIEF, CLIFORDIANISM, AND BELIEF'S AIM OF TRUTH	1
1.2 THE ETHICS OF BELIEF, JAMESIANISM, AND INDIRECT DUTIES TO BELIEVE	5
1.3 FROM THE ETHICS OF BELIEF TO THE ETHICS OF MIND	8
1.4 SUMMARY	10
<u>2 MENTAL RESPONSIBILITY AND REASONS FOR ATTITUDES</u>	<u>12</u>
2.1 HOW WE SEEM TO LACK MENTAL CONTROL	13
2.2 HOW WE SEEM TO BE DIRECTLY RESPONSIBLE FOR OUR ATTITUDES	16
2.3 HOW WE SEEM TO LACK DIRECT VOLUNTARY CONTROL OVER ATTITUDES	18
2.4 HOW THE LACK OF VOLUNTARY CONTROL IS NOT THE CORE OF THE PROBLEM	20
2.5 THE CORE OF THE PROBLEM: RESPONSIBILITY TO THE STANDARDS OF RATIONALITY	24
2.6 SOLUTIONS TO THE PROBLEM: HOW THERE COULD BE AN ETHICS OF MIND	26
2.7 SUMMARY	32
<u>3 WHAT ATTITUDES SHOULD WE HAVE?</u>	<u>35</u>
3.1 CLIFFORDIANISM/JAMESIANISM AND EVIDENTIALISM/PRAGMATISM	36
3.2 EVIDENTIALISM VS PRAGMATISM: A ROUGH SKETCH	38
3.3 'WHAT ATTITUDES SHOULD WE HAVE?' AS A STRUCTURAL QUESTION	41
3.4 UNDERSTANDING RESPONSIBILITY BY UNDERSTANDING REASONS: THE APPROACH OF THIS BOOK	44
3.5 THE RELEVANCE OF MENTAL INVOLUNTARISM	48
3.6 SUMMARY	51
<u>PART II REASONS FOR ATTITUDES, BLAMEWORTHINESS, AND CASES OF CONFLICT</u>	<u>53</u>

<i>INTRODUCTION TO PART II</i>	54
<u>4 THE ARGUMENT FROM DOXASTIC BLAMEWORTHINESS</u>	57
4.1 EVIDENCE AND BLAME	57
4.1.1 THE CLUTTER OBJECTION AGAINST EPISTEMIC NORMS	57
4.1.2 A DILEMMA FOR EPISTEMIC BLAME	60
4.1.3 AN ARGUMENT FOR DOXASTIC NIHILISM	63
4.2 SOME INITIAL WORRIES WITH THE ARGUMENT FROM DOXASTIC BLAMEWORTHINESS	67
4.2.1 A WORRY FROM CONSIDERING CHILDREN AND ANIMALS	67
4.2.2 A WORRY FROM OBJECTIVISM ABOUT 'OUGHT' AND 'REASONS'	68
4.2.3 A WORRY FROM EPISTEMIC INSTRUMENTALISM	70
4.2.4 A WORRY FROM EPISTEMIC PERMISSIVISM	71
4.3 SOME PERSPECTIVE: THE WEAKNESS OF THE ARGUMENT FROM DOXASTIC BLAMEWORTHINESS	74
4.4 CONSEQUENTIAL CONCEPTIONS OF DOXASTIC RESPONSIBILITY	77
4.5 SUMMARY AND OUTLOOK	79
<u>5 THE ARGUMENT FROM MENTAL BLAMEWORTHINESS</u>	81
5.1 OBJECT-GIVEN REASONS AND BLAME	81
5.1.1 GENERALIZING THE ARGUMENT FROM BLAMEWORTHINESS	82
5.1.2 HOW TO SUPPORT THE SECOND PREMISE: REASONS FOR ATTITUDES AND BLAMELESSNESS	83
5.1.3 REASONS FOR DESIRE AND BLAMELESSNESS	84
5.1.4 REASONS FOR INTENTION AND BLAMELESSNESS	85
5.1.5 REASONS FOR EMOTION AND BLAMELESSNESS	87
5.1.6 INTERMEDIATE SUMMARY	89
5.2 OUR CHALLENGE: UNDERSTANDING THE NORMATIVE FORCE OF RATIONALITY	90
5.3 WHAT IS NORMATIVITY?	95
5.3.1 NORMATIVE STANDARDS AND REASONS	96
5.3.2 NORMATIVE STANDARDS AND VALUE	97
5.3.3 NORMATIVE STANDARDS AND RESPONSIBILITY	98
5.3.4 CONCEPTS OF NORMATIVE REASONS	99
5.4 SUMMARY	101
<u>6 THE ARGUMENT FROM CASES OF CONFLICT</u>	103
6.1 DEONTIC TRANSMISSION BETWEEN RESPONSES	103
6.2 AN ARGUMENT AGAINST THE STRICT NORMATIVITY OF OBJECT-GIVEN REASONS	107
6.2.1 FROM ORDINARY STATES TO REASONS-RESPONSIVE ATTITUDES	107
6.2.2 THE TWO DOMAINS REPLY	108
6.2.3 RESPONDING TO THE TWO-DOMAINS REPLY	110
6.3 AN ARGUMENT AGAINST THE PRO TANTO NORMATIVITY OF OBJECT-GIVEN REASONS	112
6.4 IRRATIONALITY AND CRITICISM	114
6.4.1 CRITICISM, 'OUGHT,' AND THE VIOLATION OF REQUIREMENTS	115
6.4.2 CRITICISM, THICK CONCEPTS, AND MORAL BLAME	116
6.4.3 CRITICISM, IRRATIONALITY, AND VICE	120

6.5	SUMMARY	122
SUMMARY AND CONCLUSION OF PART II		125
PART III REASONS FOR ATTITUDES AND VALUE.....		129
<i>INTRODUCTION TO PART III</i>		<i>130</i>
7 WHAT IS PRAGMATISM ABOUT REASONS FOR ATTITUDES?.....		133
7.1	 OUGHT TO BELIEVE AND REASONS FOR ACTION.....	134
7.2	 TRADITIONAL PRAGMATISM IMPLIES MENTAL VOLUNTARISM.....	138
7.2.1	 SHAH’S ARGUMENT FOR EVIDENTIALISM AND ΦING FOR A REASON.....	139
7.2.2	 WAYS OF ΦING FOR A REASON	141
7.2.3	 DELIBERATION AND ΦING FOR A REASON	145
7.2.4	 PRAGMATISM IMPLIES VOLUNTARISM: ΦING FOR PRACTICAL REASONS IS ΦING AT WILL	147
7.3	 SUMMARY	148
8 FORMING ATTITUDES FOR PRACTICAL REASONS		150
8.1	 REACTING TO REASONS AND REACTING FOR REASONS	151
8.1.1	 REACTING TO, REACTING FOR, AND DEVIANT CAUSATION	151
8.1.2	 WHAT VOLUNTARISTS MUST, AND MUST NOT, ARGUE FOR	153
8.1.3	 WHY AIMING AT CORRECTNESS DOES NOT REFUTE VOLUNTARISM	154
8.2	 THE VOLUNTARIST’S DILEMMA.....	155
8.3	 RE-FORMULATING VOLUNTARISM?.....	158
8.4	 SUMMARY	161
9 ATTITUDES AS STATES		162
9.1	 ATTITUDES AND ACTIVITIES	162
9.2	 ACTIVE ATTITUDES AND INDIRECT CONTROL.....	168
9.3	 FREEDOM OF INTENTION AND DECISION.....	171
9.4	 GUIDANCE CONTROL OVER ACTIONS AND ATTITUDES.....	172
9.4.1	 GUIDANCE CONTROL AND PRACTICAL NORMATIVITY	173
9.4.2	 ACTIONS AND MENTAL STATES AS RESULTS OF MECHANISMS	176
9.4.3	 GUIDANCE CONTROL, VOLUNTARY CONTROL, AND DIRECT RESPONSIBILITY	180
9.5	 SUMMARY	183
SUMMARY AND CONCLUSION OF PART III		185
PART IV RESPONSIBILITY FOR ATTITUDES		187
<i>INTRODUCTION TO PART IV</i>		<i>188</i>

<u>10 INDIRECT RESPONSIBILITY AND ANSWERABILITY</u>	<u>192</u>
10.1 THE ORIGINAL PROBLEM RE-STATED	192
10.2 TWO KINDS OF (PUTATIVE) BLAMEWORTHINESS	194
10.3 DISTINCT KINDS OF (PUTATIVE) RESPONSIBILITY	197
10.4 A PROBLEM FOR ANSWERABILITY AS A FORM OF RESPONSIBILITY?	202
10.5 HOW WE ARE <i>NOT</i> BLAMEWORTHY FOR IRRATIONALITY	206
10.6 SUMMARY	210
<u>11 TWO FACES OF RESPONSIBILITY FOR ATTITUDES</u>	<u>211</u>
11.1 THE ARGUMENT FROM APOLOGY	212
11.2 THE DIAGNOSIS: TWO FACES OF BLAME	216
11.3 SUPPORTING THE DIAGNOSIS: THE RATIONALITY OF PASSIONATE BLAME	220
11.4 THE RELEVANCE OF REASONS-RESPONSIVENESS AND INDIRECT CONTROL.....	226
11.4.1 IS ONLY RESPONSIVENESS TO REASONS RELEVANT TO MENTAL RESPONSIBILITY?	226
11.4.2 IS ONLY INDIRECT CONTROL RELEVANT TO MENTAL RESPONSIBILITY?	228
11.5 A FURTHER ARGUMENT FOR THE DIAGNOSIS: MORAL LUCK AND VICTIM CRIMINALS	230
11.6 CONCLUSION: THE SOLUTION TO THE PROBLEM OF MENTAL RESPONSIBILITY	232
<u>12 RESPONSIBILITY FOR OURSELVES: SKETCHING AN ACCOUNT OF SELF-CONTROL.....</u>	<u>236</u>
12.1 THE OBJECTION FROM MENTAL AGENCY	237
12.2 GETTING A GRIP ON THE OBJECTION	239
12.3 THE BASIC IDEA: REASONING, CAUSING, AND RATIONAL DETERMINATION	243
12.4 REPLYING TO THE OBJECTION: A CONCEPTUAL ASYMMETRY	249
12.5 REMAINING ISSUES	252
12.5.1 ACTIVELY DETERMINING ONE'S ATTITUDES	253
12.5.2 CONTROLLING ONE'S RATIONALITY THROUGH REASONING	255
12.5.3 CONTROLLING ONE'S EMOTIONS THROUGH REASONING	258
12.6 SUMMARY	259
<u>CONCLUSION</u>	<u>261</u>
<u>LITERATURE</u>	<u>267</u>

Chapter 0**Introduction**

Are we responsible for the way we are? Our attitudes – our beliefs, desires, emotions, and intentions – are not actions. They are conceived of as mental *states* we are in, rather than things we actively perform. Intuitively, we are not responsible for just being in a state – at most, we are responsible for causing a state or for failing to avoid it. And yet we respond to and evaluate our attitudes in ways which are similar to the ways in which we respond to and evaluate our actions. We think that we *ought* to believe in human-induced climate change, and we even consider it to be appropriate to criticize others if they fail to believe in it. A malicious desire, like the desire for another’s suffering, can rightly provoke not only disapproval, but also resentment or indignation. An emotion like anger might turn out to be unjustified, and we might owe an apology to the person who was the target of our hostile emotion. And merely *intending* to become a better person is often already worthy of praise or credit.

These are undeniable commonplaces which seem to indicate that we are responsible not only for our actions – for what we *do* – but also for our attitudes: we are also responsible for who we *are*. Understanding our *responsibility* for our attitudes (if there really is such a responsibility), and the *norms* governing our attitudes (if there really are such norms), are the two main tasks of an ethics of mind.¹

This book explores the intuitively plausible idea that we are responsible for our attitudes only insofar as we can control them by what we *do*. According to this idea, an attitude can only be required of a person *indirectly* – i.e., by requiring them to *do* something in order to bring the required attitude about. This idea is at odds with a widespread assumption within contemporary epistemology and metaethics: that attitudes like belief and intention can be *directly* required – i.e., *without* requiring us to bring about the belief or the intention by means of an action. According to this idea, it can be required of a person to *just* believe or *just* intend something. Metaethicists talk about ‘requirements of rationality’ that are supposed to directly govern our attitudes rather than any actions of mental self-management.² In epistemology, William P.

¹ That does not mean that the ethics of mind is without subject matter if it turns out that there is no such responsibility and that there are no such norms. The ethics of mind would then be concerned with explaining why this is so, and how it can *seem* that we are responsible for the way we are and why it *seems* that our mental states are governed by norms.

² On rationality, cf. 0.2.4 below. According to the two most influential positions, rationality requires us either to have attitudes which are coherent with one another (Broome 2013), or to respond correctly to our reasons for

Alston's (1988) attack on what he called 'deontological conceptions of epistemic justification' provoked elaborated defenses of the idea that we are directly responsible for our beliefs in a similar way as we are responsible for our actions.³ Furthermore, leading theorists of responsibility, like Pamela Hieronymi or Angela Smith, argue that direct responsibility for our attitudes is something we need to *explain* rather than something we could reasonably call into doubt.⁴ Finally, discussions about epistemic and practical reasons for belief, which have some of their origin in the discussions of William K. Clifford (1877) and William James (1896), are concerned with what we ought to believe, rather than with how to manage our doxastic life.⁵

I think that the common idea in these debates – that not only our actions, but also our attitudes are normatively evaluable and are something for which we are directly responsible – is not misguided. Rather, there is a sense in which we should acknowledge direct responsibility for attitudes, and a (normative) sense in which it is true that we *ought* to believe, desire, feel, or intend certain things. Yet it is not at all obvious what this sense is. For we should not just disregard our plausible intuition that it does not make sense to require people to just *be* in states – whether these states are mental or not. Instead, I propose that we should explore this intuition in order to see in what sense, if at all, it can stand up to scrutiny.

This exploration will lead us to acknowledge that there is *a sense* of 'responsible' in which it is true that we are only indirectly responsible for our attitudes. The main claim of this book is thus that we can understand our responsibility for attitudes if we acknowledge that we are, in a sense of 'responsible,' only indirectly responsible for our attitudes, but that we are, in *another* sense of 'responsible,' also directly responsible for them. Acknowledging this ambiguity in our concept of responsibility provides us with a satisfying understanding of mental responsibility. Call this the *hybrid account of mental responsibility*. I will reach this account

these attitudes (Kiesewetter 2017; Lord 2018). Both positions take themselves to be concerned with requirements to *be* a certain way – to believe, desire, feel, intend, something – rather than with requirements to perform certain actions in order to ensure one's rationality.

³ Cf. Boyle (2013), McCormick (2015), and Hieronymi (2008) for some recent defenses.

⁴ Cf. Hieronymi 2006; 2014; ms; Smith 2005; 2015a; 2015b. the debate has some of its origin in the paper of Adams (1985), who defends the idea that attitudes can be direct objects of personal criticism or blame. Hieronymi's and Smith's accounts have their predecessors in accounts of responsibility which argue that our responsibility originates in our character – our 'quality of will' or our 'real self' (Wolf 1994). Arguably, our responsibility can originate there only if we were non-derivatively responsible for the way we are. Some contemporary theorists of responsibility even argue that we are *only* directly responsible for our attitudes, but never directly for our actions (Graham 2014).

⁵ At least this is how many proponents of the debate wish to understand their views, cf. Adler 2002; Kelly 2002; Owens 2000; Wedgwood 2002. However, especially in more recent years, pragmatists like McCormick (2015), Reisner (2009), and Rinard (2015; 2017) have questioned the strict distinction between norms of belief and norms of belief-management. I reinforce the distinction in chapter 7. For a good volume on the contemporary debate on reasons for belief, cf. Reisner/Steglich-Petersen 2011.

by considering issues surrounding reasons and rationality, on the one hand (parts II and III), and our practice of holding each other responsible for our attitudes, on the other (part IV).

In this chapter, I introduce the topic of this book and the problems it is concerned with. First, I spell out the project, its motivation, and my main claims (section 0.1). I then provide some conceptual clarifications that are central to my inquiry (section 0.2). Finally, I summarize the chapters (section 0.3).

0.1 THE PROJECT

The problem of responsibility for attitudes arises, *prima facie*, from the fact that attitudes are not themselves exercises of control – mental states are nothing we, strictly speaking, *do*. They are not *actions*. We cannot *choose* what we believe or feel. Yet according to a wide-spread assumption, we are only responsible for what is within the scope of our voluntary control. According to Descartes, for example,

only one thing in us [...] could give us good reason for esteeming ourselves, namely the exercise of our free will and the control we have over our volitions. For we can reasonably be praised or blamed only for actions that depend upon this free will (1649, art. 152).

If Descartes is right, then how can we be responsible for our attitudes?

There seems to be an easy reply to this problem ready to hand. For although we cannot control states in the way we control our own actions ('at will' or 'voluntarily'), we are nevertheless sometimes able to control states *indirectly* – i.e., by means of our actions. I cannot only control the state of my apartment by cleaning it, but also the state of my mind by, for example, carefully considering evidence, or taking some time to relax. This will have influence on my beliefs and emotions which is, to a certain degree, predictable. Since this ability to indirectly control our mind (sometimes) makes us at least as responsible for our states of mind as we are responsible for the state of our apartments, an answer to the question of how we can be responsible for our mind seems within reach:

The easy solution: We are responsible for our attitudes only in virtue of the fact that we are responsible for our actions by which we can sometimes foreseeably cause our attitudes.⁶

⁶ Note that this claim does not only state a necessary condition on responsibility, but an *explanation* or a fact that *grounds* our responsibility. I will be concerned with this stronger explanatory claim. This solution can be attributed with respect to responsibility for belief to David Hume (see the interpretation in McCormick 2015, 82-84), Price (1956), and with respect to emotions to Oakley (1992, ch. 4). As far as I can see, the only explicit defenses of the easy solution with respect to doxastic responsibility in contemporary debates are Meylan (2013;

More recently, however, this response to the problem has fallen into disrepute. According to most contemporary philosophers concerned with this problem, *The easy solution* has implausible implications. For one, it seems to present our attitudes as something *external* to us – as something we can at most influence indirectly, like ordinary objects in our environment, or like our sensations and bodily processes. This in turn would mean that we are *only* responsible for our attitudes in the way we are responsible for the state of our apartment, the behaviour of our dog, or for suffering pain after we forgot to take a pain killer. We seem to be, however, directly subject to certain evaluations for our attitudes to which we could never be subject to for the state of our apartment, our dog’s behaviour, or our pain. It is, for example, common ground in contemporary epistemology that we judge ourselves to be *epistemically justified* or *epistemically rational*⁷ depending on whether our beliefs are based on sufficient evidence, which is independent of whether we conducted a preceding investigation in a (practically) responsible manner.⁸ These kinds of evaluation seem to imply that we are more directly responsible for what we believe, desire, feel, or intend than *The easy solution* can allow for: since these are evaluations of persons for their attitudes, it can seem that our attitudes must be somehow *more active* than mere states (like the state of our apartments) in order for such evaluations to be appropriate. Thus, *The easy solution* seems to misrepresent our responsibility for our attitudes and the active nature of our attitudes. It seems as if more sophisticated theories of reasons, control, and responsibility are called for. Contemporary philosophers spell out such theories.⁹

The essence of this worry for *The easy solution* is that we hold our attitudes *for reasons*, while we do not hold other things for which we are only indirectly responsible for reasons (like

2017) and Peels (2017), though Chrisman (2008; 2016; fc) should, I think, also be read along these lines. No author known to me defended *The easy solution* with respect to responsibility for *intentions* or *decisions*. Aristotle is often interpreted as defending a historical understanding of responsibility for our *character* which could be read along the lines of *The easy solution*: Jacobs (2001) presents what he views as an Aristotelian account of responsibility for character, explaining our responsibility by indirect control.

⁷ It is common in contemporary epistemology to treat ‘rational’ and ‘justified’ as equivalent, cf., e.g., Wedgwood 2017, 8. Some older works that emphasize the equivalence of both concepts are Cohen 1984, 283; Huemer 2001, 22; Fumerton 1995, 19. For a case against this equivalence, cf. Sylvan ms. I will, for the purposes of my inquiry, treat these concepts as equivalent. I do not deny, however, that there might be theoretical purposes for the sake of which it could make sense to distinguish both concepts.

⁸ Pace Kornblith’s (1983) account of epistemic justification.

⁹ Here I think primarily of the accounts of these issues presented in recent years by Pamela Hieronymi, David Owens, and Angela Smith. The way these philosophers approach issues surrounding the ethics of mind is in line with the way I do it here, though their solutions turn out to be different than my solution. Although I will discuss some of their claims at various points in this book, their solutions are grounded in a theoretical background to which I cannot fully do justice here. Rather, my more modest hope is to put a viable alternative on the table for discussion. My position will draw on insights of all three philosophers. Yet I will also disagree with them on various issues – especially about significance of our capacity to indirectly control our attitudes for an explanation of mental responsibility, which these philosophers do not, I argue, fully appreciate.

our headaches, the behaviour of our dog, or the state of our apartment). How can our responsibility for our reasons-responsive attitudes be indirect in the same way as our responsibility for these non-reasons-responsive entities is indirect? It seems that we face a classical philosophical problem. Intuitively, it is indeed puzzling how we can be held responsible for just *being* in a state, and how *being* in a state can be directly required of a person. However, if we are not directly responsible for our attitudes, as *The easy solution* claims, then it seems that we have to treat our attitudes as on a par with brute sensations and other consequences of our actions when it comes to our responsibility for them. Maybe we would even have to deny that attitudes are held for reasons. For aren't we responsible for everything that is within the 'space of reasons'? Our intuitive commitment to *The easy solution* seems to lead us into a position that is unacceptable. But denying *The easy solution* would require us to abandon our intuitive commitment. Which way should we go?

It seems that we cannot have it both: either we are responsible for everything that can be held for reasons, or we are only responsible for what we control. For it seems that not everything that is held for reasons is controlled by us – our attitudes are not actions. However, I will argue in this book that the impression that we have to make a decision between two competing accounts of responsibility is deceptive. If we acknowledge two different forms of responsibility, we can see that we are, in a sense, responsible for all our attitudes insofar as they are responsive to reasons, and that we are, in another sense, responsible only for those attitudes over which we could exercise indirect control. Our philosophical puzzlement arises only because we assume that we deal with just one concept of responsibility. But there are two.

My strategy for evaluating *The easy solution* is to consider the status of reasons for attitudes: how, if at all, are we held responsible for compliance with our reasons for attitudes? In other words, how, if at all, are we held responsible to the standards of rationality?¹⁰ I choose this strategy because what makes us pause when considering *The easy solution* is the fact that our attitudes are held for reasons. If we were not responsible for whether we comply with our reasons for attitudes, then this main worry with *The easy solution* would be off the table. In discussing the status of reasons for attitudes, however, it will turn out that denying our responsibility to the norms of rationality would amount to a denial of the very existence of reasons for attitudes: it would amount to *mental nihilism*.¹¹ Since such a radical claim should only be

¹⁰ On rationality and reasons, cf. 0.2.4 below.

¹¹ I adopt the label from Kieseewetter (ms), who only briefly considers nihilism. His focus of discussion is on anti-normativism, i.e., the claim that our reasons for attitudes are not normative reasons (which he rejects), rather than the claim that there are no reasons for attitudes at all.

endorsed as a last resort, we are confronted with the challenge of making our direct responsibility for (non-)compliance to reasons for attitudes intelligible. Part II of this book will mainly be devoted to setting up this challenge.

In part III, I reject the proposal that we can make sense of the status of reasons for attitudes by adopting *pragmatism* about rationality, i.e., by understanding our reasons for attitudes as being provided by value. I argue that pragmatism implies voluntarism, i.e., the claim that we can form attitudes at will, and voluntarism is false (for conceptual reasons). Part III thus contains an extensive argument both against the idea that we can form attitudes at will (including the idea that we can ‘believe at will’) and against the idea that we can form attitudes for so-called ‘practical reasons’ (including ‘practical reasons for belief’). Considering pragmatism is essential for my discussion because it presents an attractive alternative way of making direct responsibility intelligible in face of the threat of mental nihilism that we faced in part II.

In part IV, I defend my own proposal for understanding our direct responsibility for complying with our reasons for attitudes. I distinguish this form of direct responsibility, which I will understand as *answerability*, from the kind of responsibility we have for our attitudes in virtue of our capacity to control them through our actions, which I will call *historical responsibility*. I argue that whether we comply with our reasons for attitudes is significant for specific kinds of relationships that can only hold between responsible beings – like friendships, romantic relationships, or any other relationship that requires attitudes like mutual trust, proper regard, and goodwill (something a robot cannot show towards us).¹² Someone’s irrationality – i.e., one’s failure to comply with one’s reasons for attitudes – can impair these relationships. The direct responsibility at issue can be understood as giving rise to forms of blame that *mark* these impairments (Hieronymi 2004). Since these forms of blame do not require that the person who is blamed for their attitude was in control of their attitude in some way, I can deny that responsibility requires control in this sense of ‘responsible.’

Yet the intuitive plausibility of *The easy solution* – deriving from the fact that it is indeed puzzling how we can be responsible for just *being* in a state – should not be wholly disregarded. So, hopefully, there is another way of doing justice to this intuitive plausibility besides endorsing mental nihilism. The way I propose is to acknowledge that *The easy solution* is true in a sense of ‘responsible’ that is different from the sense in which we are responsible

¹² These are the attitudes Strawson (1962) viewed as essential for the kinds of relationships we can enter into with other responsible beings. Note that my remark in brackets about robots is not an empirically contentious claim about artificial intelligence. It is a conceptual point. If there are robots who can be trusted (rather than merely relied on), who show regard and goodwill, then they enter with us into ordinary interpersonal relationships and are no longer ‘robots’ in the sense of the word I use it here.

for complying with our reasons for attitudes. I will understand this sense of ‘responsible’ as being a proper basis for emotions like resentment, indignation, guilt, or similar *passionate responses* that are rational responses only to the violation of *practical requirements* (i.e., requirements of prudence and morality that govern our voluntary conduct). Violating practical requirements can legitimately give rise to the reactive sentiments, like resentment, indignation, certain forms of anger, or guilt. Thus, I argue that these more passionate emotions are never appropriate responses *merely* to someone’s attitude (as opposed to someone’s actions, or to someone’s attitude that was caused by previous actions). *In this sense* of ‘responsible,’ we are never directly responsible for our attitudes. Thus, *in this sense* of ‘responsible,’ which is defined in terms of the reactive sentiments, *The easy solution* is true.

This is the summary of my *hybrid account of mental responsibility*: in one sense of ‘responsible,’ we should accept *The easy solution* and allow that mental responsibility is to be fully explained by the fact that we have indirect control over our mental states; in another sense of ‘responsible,’ we should reject *The easy solution* but at the same time acknowledge that it is not a big deal to just accept that responsibility does not require control: our answerability for our attitudes, which is sufficient to ground the kind of responsibility that is the basis for blame as a marker of impaired relationships, does not require that one is in control of one’s attitudes.

So far for a first sketch of my position. To defend it, the book engages in various debates about reasons, rationality, and responsibility. The chapter summary in 0.3 will give an overview over how I reach this account by engaging in these debates. There I will emphasize how my investigation is not only relevant for theorists of responsibility, but also for theorists of rationality and reasons. I think both fields of philosophical investigation should be understood as being deeply intertwined. However, before I turn to this summary, some of the concepts that are central to my investigation need to be clarified.

0.2 SOME INITIAL CONCEPTUAL CLARIFICATIONS

I will not provide precise definitions of each concept I discuss here. Rather, I will give the reader a first grip on how I use the concept in the investigation to come. It is probably most obvious with my concept of responsibility that clarifying the concept will be an essential part of my answer to our problem. This answer cannot be spelled out fully before the end of part IV of this book. I think this is the nature of any philosophical investigation that is purely conceptual: if we clarify all the concepts, then the investigation is over. Thus, we will have to rest content with some initial grip on the ideas of norms (0.2.1), the relation between norms and

reasons (0.2.2), responsibility and blame (0.2.3), as well as the concepts of rationality and its relation to reasons and types of reasons (most relevantly, object-given and state-given reasons) (0.2.4).

0.2.1 NORMS, STANDARDS, RULES, AND REQUIREMENTS

Norms are the most general term of the concepts mentioned in the heading of this subsection. A norm separates cases that conform to the norm from cases that do not conform to it. For example, we have norms for sizes of papers (DIN A4) by convention, presumably in order to have uniform sizes that make it easier to handle paperwork. The A4-norm separates papers that match the size defined by it from papers that do not match this size. Of course, that a paper is not A4 does not mean that there is anything *intrinsically bad* about it. We have the norm because, in certain contexts, it is helpful to have papers that are A4 (or, in other contexts, to have papers that are A1, A2, A3, etc.). In this general sense of the term, agreeing on a norm is just drawing a line between things that fulfill the norm and things that do not fulfill it.

There are other descriptive uses of the term ‘norm’ that are, like the example just given, not of primary interest for my investigation. Sometimes we say things like ‘It is the norm to be married by the age of 30,’ thereby referring to some regularity within a society. Any such purely descriptive use of the term does not imply anything about what one normatively *ought* to do. Such a sentence might also express the beliefs of people within the society about when one ought to be married. But also the beliefs of these people will not have any straightforward implications for what individuals normatively ought to do.

The A4-norm is an example of what I will call a *standard*. This is a specific type of norm that can be distinguished from *rules* and *requirements*. Standards do not presuppose that anyone has control over whether the standard is fulfilled. Even if we were all bound to chairs for all eternity from now on, the A4-standard would thereby not become invalid (though useless). We would still be able to reminisce with each other about the times where we produced papers matching the A4-standard. Such standards are conventional. There are not only standards for paper-sizes, but also for sizes of other artefacts, like cars, bottles, chairs, windows, etc. Yet there are also non-conventional standards. John Broome, while distinguishing types of ‘oughts’ or ‘shoulds’¹³ mentions ‘an oak should have deep, sturdy roots’ (2013, 12).¹⁴

¹³ Throughout this whole investigation, I will not make a difference between ‘ought’ and ‘should.’

¹⁴ Broome takes the example from Philippa Foot. He writes (same page): ‘An oak’s having deep, sturdy roots could fairly be called “a norm,” and that is enough to justify the term ‘normativity.’ However, in the context of an oak, natural normativity is not what I call true normativity. To say an oak should, in the sense of natural

Such a non-conventional standard is a standard of *normal development* of an oak. It is how we expect oaks to develop if they grow in the right environment. Such a standard is non-conventional because it is set by the natural development of an organism, rather than by human arbitrariness.

Another example of standards are the *standards of rationality*. A standard of rationality could be

(TR) One (rationally) ought to properly base one's beliefs on one's evidence.

If one does not fulfill (TR), say, because one believes something for which one lacks sufficient evidence, one would be irrational in one respect and to a certain degree (*if* (TR) is a standard of rationality). I will argue throughout this book that for such standards to be in place we need not be in control over whether we are rational. Yet for my categorization of the norms of rationality as *standards* I take it to be essential that these standards do not presuppose that we can fulfill them by exercising *voluntary* control. The presupposition of voluntary control is what distinguishes standards from rules and requirements, to which I will turn below. The assumption that the norms of rationality do not presuppose that we can choose whether we are rational is not wholly uncontroversial. I will support this assumption in chapters 2.2, 2.4, 4.4, and part III.¹⁵

In the contemporary philosophical literature, the standards of rationality are usually called 'requirements.' I think this is confusing. We are required to *do* certain things – i.e., to perform certain actions – but we are not required to just *be* in a certain way. This is something the *The easy solution* gets right. We cannot be rational 'just like that' or 'at will.' If we suffer from irrational angry episodes, for example, then we are not *required* not to be angry. And yet we are evaluated in the light of the standards of our rationality. We might also have reason to get rid of our disposition towards those angry episodes by doing therapy. But this is a requirement *to take steps* rather than a requirement to be rational 'just like that.'

The idea that the norms of rationality are requirements gives rise to the confusion that being subject to these norms requires that one is in control of whether one complies with these norms. Requirements presuppose voluntary control, but the standards of rationality do not. It

normativity, have deep, sturdy roots *is only to say it needs deep, sturdy roots to complete its life cycle*. This is not a truly normative statement.'

¹⁵ Part III argues against the idea that we can control our attitudes at will (i.e., we do not exercise direct voluntary control over our attitudes). Ch. 2.2 argues that direct voluntary control does not explain why we are responsible to the standards of rationality if they are understood in terms of object-given reasons. Ch. 2.4 and 4.4 argue that indirect voluntary control cannot explain why we are responsible to the standards of rationality.

is a requirement of morality to help others in severe need. This requirement presupposes that one is able to do so – that one can *make an effort, take steps, or set out* to help them. In the absence of any opportunity to exercise one’s capacity of voluntary control – i.e., in the absence of any reasonable course of action –, one is not subject to a requirement to help people in severe need. One’s inability excuses. Yet one can have sufficient evidence for a claim and irrationally fail to believe the claim even if there is nothing one could have done about one’s failure of rationality. The inability to ‘believe otherwise’ does not make one’s failure not irrational.¹⁶

Rules, a subtype of norms which will not play a big part in the discussion to come, but which should be mentioned nevertheless, also seem to presuppose voluntary control. Paradigm cases of rules are rules of a game. The rules of a game regulate what kinds of actions we are allowed to do if we want to play the game. They regulate our voluntary conduct, rather than our attitudes. These rules would thus lose their point if they were addressed to someone who cannot comply with them. Philosophers sometimes talk about ‘rules of inference’ or ‘rules of logic.’ This might imply that there is also such a thing as implicit, involuntary rule-following, as when we draw conclusions by using logical principles (like *modus tollens*) or solve simple math exercises without much reflection. However, I think it is up to debate whether the talk of ‘rules’ here is only derivative to the primary sense in which rules are meant to regulate voluntary conduct. If I am wrong that this use is merely derivative, then the standards of rationality might well be conceived as rules. But this will not matter for my main claims and is mainly a terminological controversy.

0.2.2 NORMS AND REASONS

What is the relationship between norms and reasons? Norms can be expressed with ‘oughts’ or ‘shoulds’ (I use the terms as equivalent throughout the book) as in ‘one ought not to steal’ (which is equivalent to ‘one is not allowed to steal’) or ‘an oak should have deep, sturdy roots.’ Yet not all norms provide us with reasons (cf. Kiesewetter 2017, 3-4).¹⁷ My latter example illustrates this: just because it is true, in some sense, that oaks should have deep, sturdy roots, it does not follow that anyone has a reason to, say, plant well-developing oaks or maintain the

¹⁶ I do not deny that there is a sense in which someone who is irrational ‘can’ be rational. My claim is merely that this ‘can’ is not to be spelled out in terms of control.

¹⁷ For another understanding of normativity in terms of reasons, cf. Raz 2011, 85. I will follow Kiesewetter and Raz in this understanding for now. However, in chapter 5.3, I will discuss the concept of normativity in more detail and argue that, since it is a philosophical artefact, there are different legitimate uses of the concept, and it is helpful to clarify in what sense we are arguing that a specific norm is normative or not. There I also present Kiesewetter’s concept of normativity in some more detail.

roots of oaks. Similarly, just because the rules of a game require to make a certain move does not imply that one has a reason to make the move (maybe there is no reason to play the game). Or, to give one last example, just because the rules of etiquette require one to use a certain fork for a specific meal does not imply that one always has a reason to use that fork for that meal (sometimes we do not care about etiquette, and sometimes nobody will be offended by not complying with etiquette).¹⁸

Those norms that provide us with reasons – requirements of morality and prudence, and (most likely) the standards of rationality – are often called *normative* requirements or, respectively, *normative* standards.¹⁹ We can, in line with the metaethical literature on reasons, define the following conceptual relations between normative standards or requirements and reasons that will be relevant in the following book (I will discuss the notion of normativity in more detail in chapter 5.3 after having related my discussion to the discussion about the ‘normativity’ of rationality in chapter 5.2):²⁰

Decisive reasons. S has decisive reasons to ϕ iff S (normatively) ought to ϕ .

Sufficient-allowed-link. If S has sufficient reason to ϕ , then S is (normatively) allowed to ϕ .

Sufficient reasons. S has sufficient reasons to ϕ iff S (normatively) ought to [ϕ or give any other response that is allowed].

Decisive reasons are not only to be contrasted with sufficient reasons, but also with *pro tanto* reasons. A *pro tanto* reason to ϕ neither needs to be sufficient nor decisive. Rather, it can sometimes be easily outweighed by other reasons even though it favours ϕ ing.²¹ For example,

¹⁸ Brössel et al. (2013) argue that every domain provides us with domain-specific reasons. For example, there are ‘financial reasons’ that we have whenever a decision is financially beneficial (cf. 2013, 285). Thus, even if we do not need any money at all (say, because we live self-sustainably somewhere far out in the forests), we would still have a *financial reason* to enrich our finances. Similarly, we would always have an ‘etiquette reason’ even though there is no independent reason to comply with the norms of etiquette in a given situation. Kiesewetter (ms) convincingly criticizes this view as merely stipulating a notion of domain relative reasons. It is much more natural to say that, if we do not need any money at all, then we also have no reason to enrich our finances. Thus, the norms of financial maximization or the norms of etiquette are paradigms of norms that do not provide us with reasons.

¹⁹ To be precise, these norms are all called *normative requirements* in the literature, partly because it is odd to talk about ‘normative norms,’ and partly because it is assumed that rationality places requirements on us. According to my distinctions above, however, rationality is a standard rather than a requirement. Note that I am not saying anything controversial here: I am not saying that the standards of rationality are ‘merely evaluative’ rather than ‘normative.’ Rather, I will argue, while assuming certain things about the concept of rationality (cf. 0.2.4 below) that they are not mere evaluative standards, but that they provide us with reasons for conformity.

²⁰ For similar definitions and some more details on sufficient and decisive reasons, cf. Kiesewetter 2017, 8-9.

²¹ However, it is not essential for a *pro tanto* reason that it can easily be outweighed. Some *pro tanto* reasons are very strong reasons for an action. In cases where an action is strongly supported by a *pro tanto* reason, but yet we are not allowed to do the action because there is a better course of action available, we usually feel some kind of regret for not being able to both do what we ought to do and comply to all our *pro tanto* reasons.

the fact that jumping on the neighbor's flowers would be lots of fun is a reason to jump on the neighbor's flowers. Yet this reason neither makes it the case that I have decisive reason to jump on the flowers – I am not required to jump on them. Nor do I have sufficient reason to do so – it would not be (morally) allowed to do so. Other reasons, like the fact that it would upset our neighbor, that the flowers are beautiful, and that I might get sued, count against jumping on the flowers. And yet the fact that jumping on the flowers is fun favours jumping on the flowers and is thus a *pro tanto* reason to jump on them.

The distinction between decisive or sufficient reasons and *pro tanto* reasons holds for reasons for attitudes as well. That there is a car driving towards my house can be – given my overall information – a reason to believe that my expected guest is arriving. However, if it is possible that another person is driving towards my house, then this reason does not make it the case that I *ought* to believe that my guest is arriving. Nor would it provide me with *sufficient* epistemic reasons for believing that my guest is arriving. I might be epistemically at fault if I give in to wishful thinking and believe that my guest is arriving just because there is a car driving towards my house. The car might just make a turn and drive back, so that it turns out that it was not my expected guest who was driving it. The fact that the car is driving towards my house is, under these circumstances, just a *pro tanto* reason to believe that I will have a guest. By contrast, if the car would not turn around but I would rather see someone waving up to me from the car, then I would have, *ceteris paribus*, decisive reasons to believe that my guest is arriving.

Having decisive reasons to ϕ means that one ought to ϕ *all-things-considered*. That is, given all the reasons that are relevant for whether to ϕ , ϕ ing is the thing to do (or the attitude to have). Some authors call this the ‘deliberative ought’ (Kiesewetter 2017, 9)²² or the ‘central ought’ (Broome 2013, 22-5)²³. The ‘ought’ of decisive reasons is to be distinguished not only from ‘oughts’ that do not provide us with reasons,²⁴ it is also to be distinguished from the domain-relative ‘oughts’ of prudence or morality: If one *prudentially* ought to take a bath, it is not thereby something one has decisive reasons to do (there might be strong moral reasons against taking a bath due to the water and energy wasted); and it is a question open to

²² Kiesewetter adopts the idea to define this ‘ought’ as the one that is central for our deliberations from Williams (1965, 124).

²³ Broome (2013, 22-5) characterizes this ‘ought’ as the ‘ought’ that figures in the idea that it is irrational not to intend what one believes one *ought* to do.

²⁴ Like standards of normal development (‘an oak should have deep, sturdy roots), mere epistemic predictions (‘the train ought to arrive in a minute’), or merely evaluative ‘oughts’ (‘the world ought to be a better place’). Evaluative ‘oughts’ can provide us with reasons if we are in a position to make them true (by making the world a better place).

philosophical dispute whether what we *morally* ought to do can be outweighed by non-moral reasons. A question that will be of high relevance for our discussion is whether what we *epistemically* ought to believe expresses a domain-relative ought, or whether we always have decisive reasons to believe what we epistemically ought to believe. Are our reasons for belief merely provided by our evidence (which, I assume, is what provides us with *epistemic* reasons), or are there any non-epistemic reasons for belief? To anticipate, I will argue (in part III) that epistemic rationality is not just one domain amongst others according to which the deontic status of beliefs is evaluated. Rather, the rationality of belief is purely a matter of epistemic reasons.

There are extensive discussions about what kinds of things reasons are,²⁵ how they are related to an agent's responses to reasons,²⁶ and what types of reasons there are.²⁷ Some of these questions will become relevant for our discussion, and I will note them when the time comes, but many can be ignored.²⁸ I just mention one commitment that is not too controversial within contemporary analytic philosophy: reasons – at least *normative* reasons – are facts. When we ϕ for a reason, then our ϕ ing can be made intelligible in light of that reason: we can explain why we ϕ by reference to the reason for which we ϕ ed. Yet in order to explain something one needs to cite a fact: something that is not the case cannot explain anything. I thus take reasons neither to be mental states (though reasons can be facts about one's mind) nor to be sometimes false state of affairs or considerations. If a state of affairs or a consideration turns out to be false, then it thereby turns out that it is not a reason. Yet false state of affairs or considerations can be *apparent reasons*. This is important because there are cases in which the same (or at least, a similar) relation holds between a false consideration and a subject's response

²⁵ For a recent helpful discussion, cf. Alvarez (2018), who defends factualism about reasons. For the classical psychologism about reasons (reasons are mental states), cf. Davidson 1963. As I explain below, I follow Alvarez.

²⁶ Here debates about the 'basing-relation' between a reason and a response for this reason are relevant. On acting for a reason, cf. Davidson (1963) as the *classicus locus* of the debate. Davidson defends the idea that the reason or which we act is the primary cause of that action. Anscombe (1957) provoked positions that deny that explaining actions by citing the agent's reasons are causal explanations. For some overview over the debate, cf. the volume of D'Oro/Sandis 2013. On the basing-relation in epistemology, cf. Korcz 2019. For a recent solution to the problem of deviant causation that is central to these debates, cf. Lord 2018, chs. 5 and 6.

²⁷ The probably most salient distinction is between normative, motivating, and explanatory reasons. The distinction between these types of reasons is not necessarily ontological, but functional: what justifies my action or attitude (normative) might be what motivates me to do it (motivating), and thus explains why I do it (explanatory). Yet these reasons can come apart in various ways. For a good overview, cf. Alvarez 2017. Another distinction is the one between reasons for actions and reasons for attitudes, which is highly relevant to my discussion, and to which I turn below in ch. 0.2.4 in discussing the distinction between state-given and object-given reasons.

²⁸ Types of reasons will be relevant for the distinction between object-given and state-given reasons for attitudes, to which I return below (0.2.4). The relation between reasons and an agent's responses to these reasons will be of central importance for my argument against pragmatism in part III.

as it holds between a reason and a response given for this reason. In such cases, the subject responds merely to an apparent reason.²⁹

0.2.3 RESPONSIBILITY AND BLAME

There are norms to which we are responsible, and norms to which we are not responsible. To be responsible to a norm means that one can, at least sometimes, be legitimately blamed or personally criticized for violating the norm. Again, the requirements of morality and prudence are norms to which we are responsible. Stealing someone's purse is something you can be resented for by the affected person, and others can legitimately feel indignant about your thievery. You can regret that you did not start out to do earlier what you really want to do in life, and in this sense blame yourself for such a prudential failure.

By contrast, we are not responsible to the rules of etiquette, or to the rules of spelling. Although it seems that we sometimes *can* be legitimately blamed for failing to use the correct fork for the meal, or for misspelling another person's name, such blame is ultimately *moral* (or maybe, in some cases, prudential) blame. What we are blamed for is not showing proper regard to people attending the dinner, or to the person whose name one misspelled. There is nothing wrong *per se* with using the other fork or writing down letters in another order than one is supposed to. The blame at issue is not 'etiquette blame' or 'spelling blame.' Rather, it is moral blame. Because of this, we can say that we are not responsible to the norms of etiquette or spelling.³⁰

The distinction between norms to which we are responsible and norms to which we are not responsible is strikingly parallel to the distinction between norms that provide us with reasons and those that do not, which I mentioned in 0.2.2 above. I think this is no coincidence. Rather, philosophers sometimes understand a norm that provide us with reasons – like an epistemic standard, or a requirement of morality and prudence – as a norm 'that someone, the subject of the norm, is accountable for conforming to (in suitable conditions)' (Kauppinen 2018, 3). I will argue throughout this book that if an attitude is responsive to reasons (in a very minimal sense of 'being answerable' that will be discussed in chapter 10) is indeed sufficient for being responsible for this attitude (in *one* sense of 'responsible').

²⁹ I here follow Alvarez 2010, cf. esp. her ch. 2.

³⁰ An objection states that whether we comply to the rules of etiquette and proper spelling is under our voluntary control. If something is under our voluntary control, then surely, so the objection goes, I am responsible for it. This is true, but it does not affect my point. We are responsible for picking a certain fork or for writing down a name in a certain way. But we are thereby not *responsible to the norms* of etiquette or proper spelling, but rather responsible to practical requirements of prudence or morality.

The idea that responsibility is a matter of potentially being an appropriate object of blame (as well as positive attitudes, such as praise or gratitude) for one's actions, attitudes, or character can be traced back to Peter Strawson's (1962) proposal that we can understand our responsibility practices by understanding the reactive attitudes involved in the kinds of relationships that only responsible beings can engage in with one another. Paradigm examples of blame are resentment and indignation as other-directed blame, and guilt as self-directed blame. These emotions are only meaningful reactions towards actions and attitudes for which we are responsible. The other-directed versions of blame often have a certain significance for our relationships with other people in that they inform the other person that we take their action or attitude to impair our relationship with them. There is a significant discussion about what other reactive attitudes are instances of blame.³¹ Does blame necessarily have an emotional or even passionate component (like the attitudes just mentioned)? (cf. Wallace 1994; 2011) Or can we also blame someone by merely distrusting the person, or even by merely *judging* (dispassionately) that our relationship towards this person is impaired due to their action or attitude? (Hieronymi 2004; ms; Smith 2013; Scanlon 2008) Maybe blame can be analyzed in terms of behavioral dispositions that are organized around a characteristic belief-desire pair? (Sher 2006; 2009).

Considering the nature of blame – and the metaphilosophical question of what this dispute amounts to – will become essential for understanding our responsibility for attitudes in part IV. For the beginning, we can help us with a provisional, negative understanding provided by Tim Scanlon, who notes that blame is neither mere evaluation of a person nor a form of punishment (2008, 122). I take this to be a consensus about the nature of blame. If I note that someone lacks good eye-sight, I evaluate the person, but I do not blame them. And if I blame them, I do not necessarily thereby punish them: blame is usually an involuntary reaction to perceived wrong or norm-violation, and can thus not be considered as punishment, which is voluntarily imposing a burden on another person (cf. Hieronymi 2004).³²

Understanding responsibility in terms of the openness to reactive attitudes, an especially blame, leaves it open what other conditions must be satisfied for someone to be blameworthy besides that one is responsible. For we are responsible for many of our actions and attitudes without being blameworthy for them. I am responsible for typing these words, but I

³¹ Cf. esp. the volume of Coates/Tognazzini 2013. See the introduction there for an overview of positions about the nature of blame.

³² Blame might be instrumentalized as punishment: I might cultivate my blaming emotions towards another person such that I thereby punish them and make them do what I want. Such instrumentalizing usually involves problematic manipulation of others.

am not blameworthy for doing this. One condition that needs to be fulfilled for me to be blameworthy is that by performing an action, I violate a norm to which I am responsible, like a requirement of morality. However, it is not clear that I am always blameworthy for violating a norm to which I am responsible. For there might be such things as *excuses*. If you forget your best friend's birthday, but only because some horrible event distracted you from remembering it, then your friend should excuse you from this interpersonal expectation to remember each other's birthdays. A promising definition of blameworthiness thus seems to be:

- S is blameworthy for ϕ ing iff
- S is responsible for ϕ ing, and
 - by ϕ ing, S violates a norm to which S is responsible, and
 - S is not excused for violating the norm by ϕ ing.

This is a definition of blameworthiness in terms of responsibility. It is circular, because I understood 'being responsible to a norm' in terms of being a potentially appropriate object of blame for violating the norm above. Yet I have already pointed out that I think that the norms to which we are responsible are the norms that provide us with reasons. The second condition can thus be rephrased as 'by ϕ ing, S violates a norm that to which S ought to comply (i.e., has decisive reasons to comply)' in order to eliminate the circularity.

According to this definition of blameworthiness above, responsibility is the property that must be in place for someone to be blameworthy: it is the most essential precondition for blameworthiness in that the person must both be responsible for what they are blameworthy for and responsible to the norm they violate. Responsibility, one might say, is the basis for all our practices involving the reactive attitudes – it is what makes these practices legitimate (in some sense).

It is common to distinguish *moral* responsibility from *causal*, *legal*, and *role* responsibility. A stone can be causally responsible for smashing a window, but the stone is neither accountable legally nor morally for the broken window, because the stone is no agent. In certain contexts, we are legally responsible without being morally responsible. We might be required by law to pay the bill for a damage we caused, although there is nothing we could have done to avoid the damage we have to pay for.³³ We might also be responsible for something in virtue

³³ Following Frankfurt (1971), some authors deny that alternative possibilities are a necessary condition on moral responsibility. I tend towards disagreeing with this, but it will not matter for my argument in this book. Whatever the conditions on moral responsibility that one accepts, they might not be fulfilled and still one would have to pay the bill because one is legally responsible.

of our role without being morally responsible. If we are the head of a company, we will have certain ‘responsibilities’ in the sense of *obligations* in virtue of that role. If we turn out to be incapable of fulfilling these responsibilities, and are thus not *morally* responsible for fulfilling them (due to the incapability), then we will still have to suffer the consequences (like losing our position).

Yet it would be misleading to talk about ‘moral responsibility’ for the property I am focusing on. As I have mentioned above, we might also be *prudentially* responsible as when we feel regret about not having done something earlier. Furthermore, a central question I will pose is whether and how we are responsible to the standards of rationality. If we are responsible to these standards, then the responsibility in question is – likely – not the same as moral responsibility. I thus use the term ‘responsibility’ as encompassing what philosophers usually understand as moral responsibility, but also including prudential responsibility, and potentially responsibility to other standards, rules, or requirements that provide us with reasons, like the standards of epistemic rationality. Distinguishing how we are responsible to different kinds of norms is a task for part IV.

0.2.4 RATIONALITY AND REASONS

To introduce the topic of rationality and how I use the term ‘rational’ as intimately connected with reasons, I will first introduce the recent debate that is relevant for my investigation. I will explain where my use of ‘rational’ connects to within this debate. I then explain the distinction between object-given reasons, state-given reasons, and practical reasons for attitudes. This distinction is of high relevance to my investigation and it will help me to specify my use of ‘rational.’ To be rational, according to my terminology, will be understood as *responding correctly to one’s object-given reasons*. It will only be throughout part III that I call this conception into question by taking pragmatism about the standards of rationality seriously as an alternative conception.

The examples I mentioned as standards do not provide us with reasons. The fact that certain papers ‘ought’ to be A4 or that oaks ‘should’ have deep sturdy roots does not imply that one always has a reason to ensure that papers are A4 or to ensure that oaks have deep sturdy roots. Sometimes it might be true that we have to mess with the size of a paper or to dig the roots of an oak out of the ground – and in these cases we might rationally do so without regret or feelings of guilt. It is different with the requirements of morality. When something is morally

required, we at least have a strong *pro tanto* reason to do it, if not even a *pro toto* or decisive reason.

What about the standards of rationality? Does the fact that we *rationally* ought to have a belief, desire, emotion, or an intention imply that we have *a reason* or that we *all-things-considered ought* to be rational? Answering this question with ‘yes’ is to defend the *normativity of rationality* (cf. Kieseewetter 2017; Lord 2018). The standard worry with the idea that rationality is normative arises from a conception of rationality as mental coherence. To be rational, according to this conception, is that one’s attitudes are coherent with one another in specific ways. Standards of rationality might thus take forms like: ‘If you believe that you have sufficient evidence for p , then you ought to believe that p ’ or ‘If you believe that you ought to ϕ , then you ought to intend to ϕ ’ or ‘If you believe that you ought to ϕ , and you believe that ψ ing is a necessary means to ϕ ing, then you ought to intend to ψ .’

One of the main worries with the normativity of these standards is that they would, if they were normative, give rise to unacceptable *boot-strapping* (Kieseewetter 2017, ch. 4; Kolodny 2005, 514-42): we can make it the case that we ought to believe or intend something just by adopting the antecedent attitudes without any proper basis for them. According to the second standard of rationality, for example, it would be true that I ought to intend to scream at you if I now just arbitrarily adopt the belief that I ought to scream at you. This seems implausible. In reply, it has been suggested that the standards of rationality take wide-scope rather than narrow-scope form.³⁴ However it has been argued convincingly that also the wide-scope versions of the standards give rise to unacceptable boot-strapping (Kieseewetter 2017, chapters 4.4-4.7). Furthermore, wide-scope standards of rationality seem to implausibly imply that each way of satisfying the standard is rationally on a par (Kieseewetter 2017, chapters 6.4-6.5).

This and other problems³⁵ for a normative conception rationality as mental coherence has provoked defenses of the idea that rationality is, contrary to first impression, not a kind of mental coherence. Rather, these accounts argue that to be rational is *to respond correctly to one’s reasons* (Kieseewetter 2017; Lord 2018).³⁶ These accounts avoid boot-strapping because

³⁴ The wide-scope versions of the standards mentioned above would be ‘you ought to [not believe that you have sufficient evidence for p or believe that p]’, ‘you ought to [not believe that you ought to ϕ or intend to ϕ]’, and ‘you ought to [not believe that you ought to ϕ or not believe that that ψ ing is a necessary means to ϕ ing or intend to ψ]’ (cf. Kieseewetter 2017, 88). What is peculiar about these standards is that they can be satisfied in more than one way – i.e., by giving up or adopting one of the attitudes mentioned in the standard.

³⁵ Another problem for the normative conception of rationality as mental coherence is to say what the reason is that rationality provides us with (cf. Kieseewetter 2017, ch. 5; Kolodny 2005, 547-551).

³⁶ Wedgwood argues that the distinction between rationality as mental coherence and rationality as responding correctly to reasons is ‘illusory’ (2017, 12). However, this will depend on what one means with ‘coherence.’ Insofar as rationality supervenes on the mental even according to reasons-responsive accounts (as I explain

in cases in which we adopt an attitude without any reason for this attitude, we would not respond correctly to all our reasons even if we then adopt all the attitudes that are coherent with the adopted attitude. For example, if I now believe that I ought to scream at you without any reason, then it does not implausibly follow, according to this conception of rationality, that I ought to adopt the intention to scream at you. This is because neither my belief nor my intention are correct responses to my reasons (the fact that I believe that I ought to scream at you is not a reason to (intend to) scream at you, nor is my belief self-justifying). Indeed, such a reasons-based conception of rationality immediately seems to imply that we always ought to be rational. For it seems to be trivial that we always ought to respond correctly to our reasons. We will see throughout part II of this book that this claim is not as trivial as it seems. Part II thus poses a *new challenge* to the normativity of rationality.

The notion of ‘responding correctly’ to reasons means both to give the response that is favoured by the reasons (for example, if one has decisive epistemic reasons to believe that *p*, then the correct response to these reasons is to believe that *p* rather than to believe not-*p* or to suspend judgment) *and* to give the response *for* those reasons. I might believe what my reasons support but my belief need not be related to my reasons in the right way so as to count as a *correct* response. If you tell me that you are in town right now I might have decisive reasons to believe that you are in town. But if I then believe that you are in town because I flipped a coin and it showed heads, then I did not respond correctly to my reasons and I am not rational. I will not be concerned in this book with spelling out what it is to believe something for a reason – although the relevant connection between a reason and a response will play a significant role in my argument of part III.³⁷

Accounts of rationality as responding correctly to one’s reasons also include the idea that it is ‘one’s’ reasons one needs to respond to correctly in order to respond the way one rationally ought to. That is, the reasons are not just any reasons there are, but reasons which are in some way *accessible to*, or *possessed by*, the subject. This is required for these accounts because they need to capture the intuition that rationality *supervenes on the mental*: whether one is rational cannot be a matter of facts that are totally inaccessible to us. Even if there *is* a

below), they are also just concerned with coherence amongst the reasons accessible to the subject and the attitudes they form. This might also be called ‘mental coherence.’ Yet it is a different matter whether you are concerned with the coherence amongst your attitudes or the coherence amongst your attitudes *and* your accessible reasons (to which you need not take any attitude). Thus, the distinction has substance, and is of central importance to the accounts of Kieseewetter (2017) and Lord (2018).

³⁷ For a recent account of the connection between a reason and a response for this reason, cf. Lord 2018, chs. 5, 6.

reason to leave my office now because it will all blow up in five minutes, this reason does not make it rational for me to leave my office if I have no clue at all about the danger I am in. Proponents of reasons-based accounts of rationality thus spell out theories according to which it is the agent's *accessible* or *possessed* reasons that determine what responses the agent ought to give.³⁸ They thereby position themselves against objectivism about reasons.³⁹

When I use the term 'rational' in the rest of this book, I assume a reasons-based conception of rationality à la Kieseewetter (2017) and Lord (2018), rather than a conception of rationality as mental coherence à la Broome (2013). To be rational is to respond correctly to one's overall accessible reasons for an attitude. Yet if it would turn out that rationality cannot be understood as responding correctly to reasons, or if it would turn out that what we *ought* to do is to respond to those reasons that there are *objectively* (rather than the ones that are accessible to us or which we possess), then my overall argument is not affected. This is because what I will be interested in is the normative force of reasons for attitudes rather than the normative force of attitudinal rationality (however understood).⁴⁰ If you do not agree with reasons-based conceptions of rationality, you are free to read my use of 'rational' as stipulative: it refers to the property of an attitude of being adequately based on the subject's accessible reasons for this attitude.

What I call the 'normative force of reasons' above will be illuminated in terms of the reactive attitudes we show if someone fails to respond correctly to these reasons (esp. in part IV of the book). A working hypothesis for the discussion to come is that one can only be praised and blamed for responding or failing to respond to reasons that are *accessible* to one. That is, if someone is fully rational – i.e., responds correctly to all the reasons accessible to one –, then one cannot be blameworthy. To be blameworthy presupposes that one fails in light of one's *own* reasons. This assumption might be questioned, and I will discuss how it might be questioned and what implications that would have in chapter 10.3. It will turn out that questioning the assumption does not pose an overall problem for my account of the normative force of reasons for attitudes, and thus to my solution of the problem of mental responsibility.⁴¹

³⁸ Cf. Kieseewetter 2017, ch. 8; Lord 2018, ch. 8.

³⁹ On objectivism about reasons and 'oughts,' cf. ch. 4.2.2.

⁴⁰ I return to the possibility that reasons for attitudes exert their normative force on us whether we know of them or not, and the consequences this would have for my argument, in ch. 10.3.

⁴¹ How understanding the normative force of reasons for attitudes is connected to understanding mental responsibility will be the topic of chs. 2 and 3 of this book.

To be rational means, according to my terminology, to respond correctly to one's *object-given reasons* for the attitude.⁴² Object-given reasons for an attitude are reasons that indicate (or constitute) facts about the attitude's object rather than about the attitude itself – facts that support the attitude, or make it rational to have the attitude, by indicating (or constituting) that the attitude (partly) fulfills its constitutive aim. For example, object-given reasons for *beliefs* are (or are provided by) evidence, because evidence indicates the truth of the object of the belief, i.e., the truth of the belief's propositional content. Scientific reports on climate change are thus object-given reasons for belief: they indicate that human-induced climate change takes place. By contrast, that I feel less existential angst if I do not believe in climate change is a state-given reason not to believe in climate change. It is not an object-given reason (i.e., no evidence) against climate-change, because this fact does not indicate that the belief fulfills its constitutive aim of truth.

Analogously, an object-given reason for a *desire* shows the object of the desire to be desirable in some respect; object-given reasons for *fear* indicate the danger of what you fear, and thus make it rational to experience fear;⁴³ object-given reasons for *intention* are reasons for the object of intention – i.e., the intended action: that I will get poisoned if I drink a toxin is an object-given reason not to intend to drink it in virtue of being a reason not to drink it; and that I get a lot of money for intending to drink a toxin is a state-given reason to intend to drink the toxin.⁴⁴

In contrast to object-given reasons, state-given reasons indicate (or constitute) facts about the state of the attitude itself. The most important category of state-given reasons under discussion are *practical reasons* for attitudes, i.e., reasons that support the attitude by indicating the attitude's value.⁴⁵ According to those who believe in such practical reasons – i.e., *pragmatists*, the fact that it would be beneficial to believe in a guarding angel (e.g., because one would

⁴² For the distinction between object-given and state-given reasons, cf. Parfit 2001, 21-22. Some refer to these reasons as 'reasons of the right kind.' However, this label might be confusing because it has its origin in a specific debate about fitting-attitude accounts of value (cf. Rabinowicz/Rønnow-Rasmussen 2004 for the wrong kind of reasons problem for these accounts; cf. the discussion about the distinction between right and wrong kind of reasons, or state- and object-given reasons, in Gertken/Kiesewetter 2017). It is disputable whether the 'state-given'/'object-given' labels are much better. I assume here that, whatever the label, we can make sense of this distinction in some way. See Heuer (2018) for some recent criticism about the distinction when applied to intentions.

⁴³ Some might prefer not to talk about the 'rationality' of an emotion. I do not have any trouble calling emotions rational or irrational. If you do, then you can call the relevant normative category 'appropriateness' or 'fittingness' rather than rationality.

⁴⁴ The reason to drink the toxin in Kavka's toxin puzzle (1983) is state-given in this sense.

⁴⁵ Kolodny (2005, 551) famously pointed out that facts about an attitude's coherence would be state-given reasons for the attitude. This would be an example of a state-given reason that would be no practical reason.

sleep better at night) could be a practical reason to believe in a guarding angel (even in the absence of sufficient evidence).⁴⁶ Analogously, facts about the value of any other attitude are considered to be reasons for this attitude by pragmatists.

It is often argued that practical reasons are no *reasons* for attitudes, but rather mere *considerations* indicating the attitude's value.⁴⁷ It is argued that they are no reasons because if they were, then we would sometimes be able to form attitudes *for* practical reasons: we could believe, desire, feel, or intend at will. We could just decide not to fear a dangerous tiger or just decide to intend to drink a poisonous substance merely because we regard fearing or intending to be beneficial. I will argue in part III that we cannot adopt attitudes at will, and that this is why practical reasons are indeed no reasons for the attitude they seem to favour.⁴⁸ I thus provide my justification for restricting the reasons relevant to the rationality of attitudes to object-given reasons not before part III.

0.3 SUMMARY OF CHAPTERS

This book aims at answering what I will call the *two main questions of an ethics of mind*:

1. How can we be responsible for our attitudes?
2. What attitudes should we have?⁴⁹

In the first part of the book (chapters 1-3) I clarify these questions and provide a framework for answering them. To get an initial grip on the topics that I deem relevant for discussing both questions, I start with a brief historical excursion to the ethics of belief of William K. Clifford and William James in chapter 1. I argue that we can observe an interesting shift of focus in philosophical thought about the norms of belief. While Clifford and James, the two main historical references in the ethics of belief, understood doxastic duties as indirect duties – i.e.,

⁴⁶ Pragmatists will disagree about when the value of a belief provides a practical reason for this belief. Moderate pragmatists, like McCormick (2015), would argue that it is not a reason for this belief if all our evidence speaks against the existence of guarding angels. More radical pragmatists, like Rinard (2015; 2017), argue that every consideration indicating the value of an attitude is a practical reason for this attitude.

⁴⁷ Cf. esp. Shah (2006). This is what evidentialists claim, two which I return in chs. 1 and 3. State-given considerations can of course be evidence and thus epistemic reasons *to believe that a certain mental state M is valuable*. But they are not thereby reasons *for M*, but rather for the belief that M is valuable.

⁴⁸ Recent pragmatists dispute that believing at will is necessary for practical reasons for belief (McCormick 2015, 28-29; Reisner 2009, 68-70; Rinard 2015). I turn to their argument in ch. 7.

⁴⁹ I also frame the ethics of mind in this way in Schmidt (fca). I will understand the second question here as a structural question (cf. ch. 3): what is the general structure of the standards or requirements that govern our attitudes? One could go into more detail and spell out *specific* mental standards or requirements. But this is a mammoth task I cannot undertake here. For discussions about more specific mental norms, see esp. the second main part of the volume Schmidt/Ernst fc.

duties to manage one's beliefs –, contemporary debates are mostly concerned with norms that apply *directly* to our beliefs, rather than norms that apply, first and foremost, to our doxastic managing practices. The distinction between direct and indirect norms of belief will help us understand what the contemporary debate between evidentialism and pragmatism is (or should be) about: it should be about the norms that directly govern our beliefs, rather than about our doxastic managing practices. I furthermore explain how the discussion in the ethics of belief can fruitfully be extended to a discussion in the ethics of mind. However, since both issues – the distinction between direct and indirect norms of belief and the ethics of mind – will be central also for the two chapters to come, this first chapter could be skipped without further ado. Its purpose is mainly introductory.

Chapter 2 then explains why a book that starts off with thinking about responsibility for attitudes ends up talking so much about reasons and rationality. There I spell out the problem of mental responsibility in more detail. I argue that the problem of mental responsibility is, ultimately, a problem about understanding how we are responsible for our (non-)compliance to reasons for attitudes, i.e., how we are responsible to the standards of rationality.⁵⁰ Chapter 2 is thus concerned with understanding the first main question of an ethics of mind. Understanding mental responsibility will require us to understand the status of reasons for attitudes. More specifically, it will require us to understand the normative force of these reasons – i.e., why we legitimately conceive of reasons for attitudes as *reasons* – by thinking about how it can be appropriate to blame and criticize us for failing to respond correctly to our reasons for attitudes. Thus, I argue that thinking about responsibility for attitudes naturally leads us to thinking about reasons and 'oughts,' and thus to the second main question of an ethics of mind. I will illustrate throughout this book how our understanding of mental responsibility can be enriched by viewing both main questions of the ethics of mind as being deeply intertwined.

Chapter 3 then entangles the second main question of an ethics of mind. I argue that there are two main readings of the question of what attitudes we should have, building on the distinction between indirect doxastic requirements and direct doxastic standards that I introduced in chapter 1. An indirect mental requirement is a norm that requires us to deal in specific ways with our mind by performing actions by means of which we can influence and control our attitudes. Our responsibility to indirect mental requirements can be fully explained by *The easy solution*. I present a pragmatic account of indirect mental requirements in chapter 7.1. Yet the main focus of this book is the status of *direct mental standards*: standards of rationality that

⁵⁰ Remember that I adopt a reasons-based conception of rationality (cf. ch. 0.2.4).

directly govern our beliefs, desires, emotions, or intentions. I provide a short introduction into the contemporary debate between evidentialism and pragmatism. Furthermore, I explain the general approach of this book. This approach consists in focusing on understanding our mental responsibility by focusing on the status of the standards of rationality. That is, we can gain insights about how we are responsible for the (non-)compliance with reasons for attitudes by understanding the status of these reasons and the standards that provide us with these reasons. Finally, chapter 3 considers and defends the significance of the answer I will give to the problem of mental normativity.

Thus, part I of this book has the overall task of spelling out what an ethics of mind is concerned with, how I propose to understand its main questions, and how I approach them. Understanding our questions properly is almost half the way to their solution.

One might wonder how the question of what attitudes we ‘should’ have is still open if I adopt a conception of rationality as responding correctly to one’s reasons (cf. chapter 0.2.4). Is it not obvious that we should always respond correctly to our reasons? In what way do we still lack an understanding of the rational ‘ought’ if we agree that being rational is responding correctly to reasons? Part II of this book will be concerned with spelling out this new challenge for a conception of rationality as responding correctly to reasons in more detail, thereby gaining a better understanding of the status of the standards of rationality and making one step towards understanding our responsibility to these standards. This new challenge arises from the fact that responding correctly to reasons is understood as responding correctly to *object-given* reasons, and that the normative force of object-given reasons, and thus their very status as *reasons*, can seem dubious.⁵¹

This *threat from mental nihilism* calls into doubt the very existence of the standards of rationality. If there are no reasons for attitudes, then there is no point in standards that say that we ought to respond correctly to these non-existent reasons. The threat from mental nihilism takes the form of two main arguments: *the argument from blameworthiness* and *the argument from cases of conflict*. The first argument will be the topic of chapters 4 and 5. The second argument is topic of chapter 6. In a ‘Summary and Conclusion of Part II’ that follows chapter 6, I re-formulate sound versions of both arguments which establish an ambiguity in our concepts of ‘ought’ and ‘reason’ (rather than the non-existence of mental ‘oughts’ and reasons for attitudes). This ambiguity paves the ground for my distinction between two corresponding

⁵¹ The challenge was, in a related version, recently recognized by Kiesewetter (ms). I return to it in ch. 5.2.

forms of blameworthiness (that arise when we violate the respective ‘ought’) and responsibility.

Chapter 4 is purely epistemological in that it focusses on the attitude of belief and standards of epistemic rationality. It has several aims. First, it spells out *the threat from doxastic nihilism*, i.e., an initially plausible argument for the idea that there are no reasons for belief at all. I begin by discussing an initially plausible formulation of an epistemic standard: if you have sufficient evidence for believing that *p*, then you ought to believe that *p*. This standard faces the familiar clutter-objection formulated by Gilbert Harman, and it is quite uncontroversial in contemporary epistemology that the standard needs to be modified in response to this objection and similar worries. However, it is not easy to see how to formulate a version of a standard of epistemic rationality which is both *epistemic* and which is a standard the compliance to which *matters* to us. By arguing that we only quarrel with our initial formulation of the epistemic standard because we implicitly assume that non-compliance with an epistemic standard should at least sometimes make us blameworthy (in some sense), I establish a connection between reasons and blameworthiness that is central to my overall account. Yet this connection also gives rise to the threat from doxastic nihilism: if this connection holds, *and* if we are indeed never blameworthy for purely epistemic failures, then we would have to conclude that there are no reasons for belief.⁵² This is the argument from *doxastic* blameworthiness.

After I have spelled this argument, I discuss (still in chapter 4) some initial worries with the connection between reasons and blameworthiness presupposed by the argument. The worries come, first, from the idea that individuals that are not fully responsible, and thus not fully blameworthy, for their conduct, can still respond to reasons; secondly, from objectivism about reasons; thirdly, from epistemic instrumentalism; and, fourthly, from epistemic permissivism. I argue that all worries are unjustified. Next, I provide an initial account of my own reaction to the threat from doxastic nihilism, which is not a worry with the connection between reasons and blameworthiness (which I endorse), but rather a worry with the idea that we are not blameworthy for purely epistemic failures. I argue that it seems that we *are* blameworthy for epistemic failures by providing a paradigm example of someone who is epistemically blameworthy, and by analyzing how we might rationally react to this person, namely, by epistemically distrusting them. Finally, I argue against indirect doxastic voluntarism – i.e., the claim that our

⁵² Remember that I assume throughout part II that epistemic reasons are the only reasons for belief. Another conclusion to draw here – instead of nihilism about reasons for belief – would be pragmatism about reasons for belief: that not all reasons for belief are epistemic, because some are practical. However, pragmatism will not be considered as an option before part III.

responsibility for belief can be explained by indirect voluntary control over belief. I argue that these accounts are committed to denying the existence of epistemic blame and thereby, assuming the connection between blameworthiness and reasons, cannot do justice to the phenomenon of epistemic rationality.

Chapter 5 brings us from epistemology into metaethics. It extends the argument from chapter 3 to *all* mental states, thereby spelling out the threat from *mental nihilism*. I call the argument establishing this threat the *argument from mental blameworthiness*. As we will see, there are some significant *prima facie* worries with the argument when we apply it to desire, intention, and various emotions. I argue that these *prima facie* worries can be met for desire and intention, but not for all kinds of emotions. Thus, considering reasons for emotions gives us another reason to doubt that the argument from mental blameworthiness goes through – next to my initial plea for the existence of epistemic blame sketched in the last chapter. The chapter then spells out how the argument from blameworthiness also provides a new *prima facie* challenge to the normativity of rationality. I end the chapter by entangling different concepts of normativity and explaining in which senses the normativity of rationality is at stake.

Chapter 6 reinforces our threat from mental nihilism, and thus our challenge for the normativity of rationality, by spelling out the *argument from cases of conflict*. In cases of conflict, we ought to cause ourselves not to respond correctly to our object-given reasons. That is, we ought to cause ourselves to have beliefs, desires, emotions, or intentions that are irrational. If we always ought to be rational, then how can there be such cases? I argue that we can construct a powerful argument from these cases to the conclusion that we do not even have *a reason* to be rational. If we do not have such a reason, then this can only be because object-given considerations lack the normative force of reasons. The mental nihilist endorses this conclusion and claims that we should deny the status of object-given considerations as reasons. The argument from cases of conflict provides me with some opportunity to spell out my general strategy of how we can meet the threat from mental nihilism. I compare the criticizability that is often involved in being irrational with the criticizability that is involved in being vicious. We can grant, I argue, that we are not always criticizable or blameworthy in cases of conflict and yet by being irrational we are in a way we (normatively) ought not to be. Although there is a conceptual connection between blameworthiness/criticizability and reasons, this connection is not so strict as to not allow for being excused for one's norm-violation in cases of conflict.

Part II thus ends with a puzzle and some hope. The puzzle is that the status of object-given reasons as reasons is questionable in the face of the argument from blameworthiness and the argument from cases of conflict. If these arguments were sound, then we would not be

responsible for whether we comply with these reasons – for then these reasons would not exist. That is, we would not be responsible to the standards of rationality because there would be no such standards in any normative sense of ‘standard.’ Yet this conclusion is quite radical, and we should not endorse it without exploring other options. Here our hope comes into play. I have sketched replies to both arguments already in the respective chapters. I will build on my initial replies and argue at length in part IV that we have misconstrued the normative force of object-given reasons by thinking of blame and criticism too narrowly. A broader concept of blame will allow us to understand the normative force of object-given reasons in terms of how non-compliance with these reasons impairs specific kinds of relationships.

Before turning to this account of blame and responsibility, however, we will, in part III, consider another plausible reaction to the arguments from part II: mental pragmatism. Mental pragmatism endorses the conclusions of the arguments in that it grants that object-given considerations lack their status as reasons. But pragmatists will argue that I misconstrued the concept of rationality from the very beginning. Being rational, according to pragmatism, is not (merely) responding correctly to one’s object-given reasons for attitudes, but (also) to one’s practical reasons for attitudes. We can thus still save the idea that there are *reasons* for attitudes by accepting that there are practical reasons for attitudes. Part III argues that pragmatism fails because there are no such practical reasons. I do so especially by building on discussions in philosophy of action.

In chapter 7, I begin by distinguishing two positions that could be called ‘pragmatism.’ The first is an analysis of sentences of the form ‘S ought to believe that p / desire X / feel emotion E / intend to perform action A.’ This version of pragmatism claims that the truth of these sentences is partly a matter of practical reasons *for actions* of mental self-management. I argue that this version of pragmatism is uncontroversial, but that it is, consequently, not denied by pragmatism’s opponents. What opponents of pragmatism deny is that we can ever believe, or form other attitudes, for practical reasons. They argue that our inability to believe for practical reasons undermines the status of practical considerations as reasons for belief. I support this opposition to pragmatism by arguing that believing for practical reasons requires that one can believe at will – i.e., that one can *directly* believe for practical reasons (that is, without any mediating action between the reason and the attitude). I do so by discussing recent arguments by pragmatists who claim that believing for practical reasons does not require believing at will. My discussion here focusses on belief, but can easily be applied to attitudes more generally.

Chapter 8 then argues that we cannot form attitudes at will. Together with my argument in chapter 7, this refutes the position that there are practical reasons for attitudes. My argument

builds on a distinction between two senses of ‘forming an attitude’ and argues that the proponent of a mental will is in a dilemma: depending on which sense we choose, we can either form attitudes directly, but not actively; or we form them actively, but not directly. Since forming attitudes at will requires that we can form attitudes both actively and directly, the idea of a mental will proves to be incoherent. Yet my argument will have a leak: it relies on the notion of ‘forming an attitude.’ Mental voluntarists can formulate their position without invoking the idea of attitude-*formation*. They can say that their claim is not that we can *form* attitudes at will, but rather that we can just *believe, desire, feel, or intend* at will in a similar way as we can *raise our hand* at will.

Chapter 9 takes this leak of my argument seriously and engages in some *mental ontology*: I argue that attitudes belong to the category of states rather than activities. The claim that we can believe, desire, feel, or intend at will thus cannot be justified by the idea that attitudes are ontologically similar to actions. I then argue that, if attitudes and actions are *not* ontologically similar, then they cannot be controlled in the same way. I illustrate this by arguing against McCormick’s (2015) strategy of establishing direct control over beliefs by defending the idea that both beliefs and actions are under our ‘guidance control.’ I argue that if attitudes and actions are ontologically dissimilar, then McCormick’s approach does not underwrite the idea that both are controlled in the same way, *even though* the idea of guidance control can be applied to both action and belief. I finally argue that the applicability of guidance control to attitudes might pose a problem for Fischer and Ravizza’s (1998) theory of responsibility for actions: in order to explain why we are responsible to practical requirements, we need to assume that our actions are not only under our guidance control, but also under our voluntary control. Guidance control is thus not sufficient for subjection to practical norms – and thus can also not underwrite pragmatism.

This concludes my argument against pragmatism in part III. Pragmatism about reasons for attitudes cannot underwrite the idea that rationality matters in such a way that we are responsible to the norms of rationality. In part IV, we will thus be left with the task of understanding the normative force of object-given reasons in order to preserve the idea that rationality matters, and that we are thus directly responsible to its standards.

I should already note at this point that I will not provide a *full* account of how being rational – how responding correctly to object-given reasons – matters. Giving such a full account would require me to engage in more details in debates about the nature of epistemic blame and other forms of blame for rational failures. This is a different project. My project is rather to provide what I will call an *outline* of such an account. This outline will be sufficient

for seeing how our responsibility to the standards of rationality differs from our responsibility for managing our attitudes. Seeing this difference allows us to resolve our philosophical puzzlement about how we can be responsible for our attitudes.

In chapter 10, I distinguish between two forms of responsibility which I argue are distinct in nature: direct answerability for attitudes and historical responsibility for attitudes. They are distinct because the forms of blameworthiness they can give rise to result from the violation of different kinds of norms, or from non-compliance to different kinds of reasons. This difference between the two faces of responsibility makes it doubtful that answerability is a form of responsibility: it presents the threat from mental nihilism in a new shape. I begin my reply to the objection that answerability is no responsibility by considering in which way this objection is right: there is indeed *a sense* of ‘responsibility’ in which answerability is not responsibility.

Yet this should not prevent us from acknowledging another sense of ‘responsibility’ in which it is important to acknowledge answerability as responsibility. Chapter 11 defends direct answerability as a form of responsibility. I spell out the *argument from apology*: it is sometimes appropriate to apologize for completely involuntary failures, and this is only explainable by assuming that we are directly responsible for our attitudes in virtue of being directly answerable for them. I spell out how the two faces of responsibility distinguished in chapter 10 are connected to different sets of reactive attitudes, and I focus on the negative blaming-responses that are appropriate towards those who are responsible in one sense, but not in the other. This will both give more substance to my distinction between indirect mental requirements and standards of rationality (a distinction which is an implication of the distinction between two senses of ‘ought’ and ‘reason’ at the end of part II), as well as allow me to formulate my *hybrid account of mental responsibility: The easy solution*⁵³ is true with respect to historical responsibility, but it is false with respect to direct answerability. Yet in order to understand direct answerability, we merely need to acknowledge that our attitudes are directly responsive to reasons which have a specific normative force – i.e., which are conceptually connected to a specific set of reactive attitudes. That is, we do not need to assume that we have any form of control over our attitudes that goes beyond their responsiveness to reasons. I summarize my reply to the problem of mental responsibility at the end of this chapter.

We thus make an excursion into reasons just to see that we have to go back to a discussion about the concept of responsibility. This is not a vicious circle, but rather an example of a necessary feature of every philosophical investigation: by spelling out connections between the

⁵³ Cf. ch. 0.1.

concepts relevant to our problem, we provide an overview (*Überblick*) of the conceptual landscape that gives rise to our philosophical perturbation. We can dissolve our original philosophical problem by, first, getting into view how our problem of mental responsibility is a problem about the status of object-given reasons (part I), seeing how we need to make the status of object-given reasons intelligible in terms of responsibility and blame (part II), and finally understanding how we cannot (part III), and how we can (part IV) make this status intelligible.

In a final chapter 12 I will make my conception of mental agency underlying my account more plausible. I argue that the distinctive first-personal perspective of an agent on their own mind can be captured by considering the way we can control our own mind through the activities of deliberation and reasoning. By thinking about what is true, desirable, or right, we can thereby commit ourselves to beliefs, emotions, desires, and intentions in a way we could never commit others to their mental states, even if we could directly cause mental states in them. Nevertheless, this kind of control does not amount to *direct* control.⁵⁴ The idea that our own mind is controlled by us in a way that is more direct and more reliable than any way in which we control the rest of the world is an outdated conception within philosophy, which was predominant in Western thought, beginning in the Stoic tradition and living on throughout Cartesian conceptions of the mind which turned out to be incoherent thanks to criticisms like the one of Gilbert Ryle (1949). We should, I suggest, throw out the baby with the bathwater and reject the idea of direct control over our mind together with the idea of our self as ‘a ghost in the machine.’

Thus, parts II and III engage in debates about reasons and control, while part IV engages in debates about responsibility and control. By approaching the issues from two angles, I hope both to inspire philosophical exchange between the debates – for they are obviously very relevant to one another –, as well as to address (and hopefully convince) a bigger philosophical audience.

That much to say about the project ahead of us. Let me now introduce the ethics of mind by critically examining a classical debate that led me to think about these issues: the ethics of belief.

⁵⁴ This way of controlling our attitudes might remind the reader of the concept spelled out by Pamela Hieronymi in a series of articles (2006; 2008; 2014) and in a recent book manuscript (ms). For Hieronymi, however, it is essential that the control at issue is distinct from *indirect* control over our mind (what she calls ‘manipulative control’). By contrast, all control that I claim we have over our mind would fall under Hieronymi’s label of ‘manipulative control’ (though I deem this notion not to be adequate in each and every case of controlling our mind).

Part I

The Project and The Problem

Chapter 1**Some Historical Background**

The present chapter traces debates about mental responsibility and reasons for attitudes back to some of its origin: the ethics of belief of William K. Clifford and William James. I first discuss the question of what we ought to believe, thereby carving out a contrast between how this question was understood by Clifford and James and how it is (or should be) understood in contemporary debates (1.1 and 1.2). I end by introducing the ethics of belief more generally and explain how discussions in the ethics of belief naturally give way to an ethics of mind (1.3). Since I return to the question of what attitudes we ought to have in chapter 3, and since the main questions of an ethics of mind are spelled out in detail in the rest of part I, this chapter can be skipped without much of a loss. Its purpose is mainly introductory.

1.1 THE ETHICS OF BELIEF, CLIFFORDIANISM, AND BELIEF'S AIM OF TRUTH

In the second half of the 19th century, Clifford and James both discussed the question of what we ought to believe. Their discussions focused on whether only our *evidence* should determine what we ought to believe. That is, are only considerations bearing on the *truth* of a proposition relevant for whether we ought to believe the proposition? Or is it rather sometimes allowed to believe something that is not sufficiently supported by the evidence in order to gain some practical good, like the tranquility a certain unjustified belief might give us? Especially in the case of religious belief, James argued against Clifford, we may lose something highly valuable by withholding our judgment as to whether God exists. According to James, even if our evidence is inconclusive, we should nevertheless believe in God because of the personal *utility* of the belief.

Clifford, by contrast, defended the following strong principle: '[I]t is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence' (Clifford 1877, 77). It is important to note that what he took to justify this (strong) claim were *moral* reasons. He argued that if we do *not* follow our evidence, and even if we do this just *once*, we thereby spoil our (epistemic) character, we put ourselves in danger of acting wrongly as well as of spreading irrationality by transmitting irrational beliefs to others. According to Clifford, since what we believe influences how we act, and since our beliefs influence each other, just *one*

belief that is not based on sufficient evidence, i.e. *one* irrational belief, is likely to cause big harm:

No real belief, however trifling and fragmentary it may seem, is ever truly insignificant; it prepares us to receive more of its like, confirms those which resembled it before, and weakens others; and so gradually it lays a stealthy train in our inmost thoughts, which may some day explode into overt action, and leave its stamp upon our character for ever. [...] And no one man's belief is in any case a private matter which concerns himself alone. [...] Every hard-worked wife of an artisan may transmit to her children beliefs which shall knit society together, or rend it in pieces (Clifford 1877, 73/75).

It is highly doubtful whether Clifford's strategy in arguing for evidentialism by giving *moral* reasons for being epistemically rational can be successful.¹ Recently, especially Allan Hazlett (2013) has argued that false beliefs can be systematically useful. Since believing upon sufficient evidence – that is, believing rationally – makes it more likely to gain *true* beliefs, believing upon insufficient evidence – that is, believing irrationally – will usually ensure the acquisition of many *false* beliefs. Thus, if we know that, *with respect to a certain topic*, having false beliefs is better than having true beliefs, it would make sense to cultivate a disposition towards irrationality *with respect to this topic*. One might be worried that one thereby spoils one's overall epistemic character, as Clifford predicts. However, Hazlett also argues that some beliefs or belief-subsystems can be isolated from the rest of our belief-system (2013, 61). We could have irrational beliefs about some topics without being irrational in general. To get some idea of how a restricted tendency towards irrationality might be useful, think, for example, of beliefs about your own capabilities, beliefs about your near and dear, or of beliefs which represent the world as meaningful, including religious beliefs.² Many of those beliefs are not backed up by sufficient evidence, but may still be useful in that they enhance our self-confidence, the quality of our relationships, or our sense of a meaningful life. For example, if you have some irrational beliefs about your partner which enable you to view them in a better light than unbiased people can view them, this might, first, make a better relationship,³ and, secondly, need not imply that you are also irrational in general when you are making judgments about other people. That is, it need not imply that your (epistemic) character is generally flawed. Clifford, however, seems to rely on the false *empirical* assumption that one instance of irrationality always causes more irrationality which then always outweighs the benefits of the first instance of irrationality.

¹ I spelled out my doubts in a bit more detail in Schmidt 2017.

² McCormick gives examples of these in the context of her argument for doxastic pragmatism (2015, 53-65). She argues there, however, that religious beliefs with a very *concrete* content cannot be justified pragmatically.

³ For some worries with practical reasons for belief stemming from the nature of relationships, cf. Crawford *fc*.

Without this premise, his strong claim that we should *always* follow the evidence because it is morally best to do so lacks any rationale.

Another problem for Clifford's view is that there are cases where irrational beliefs cannot be transmitted, and thus cannot be morally problematic (Papineau 2013, 67-8). We can conceive of someone who lives alone on an island, so that there is no danger that this person transmits their irrational beliefs or practices of forming irrational beliefs to anyone else.⁴ Here it seems that we can conceive of a case in which an irrational belief does not do any harm, but can rather be quite useful for the person having the belief.⁵ Imagine, for example, that an eremite irrationally believes that she can build houses from trees faster than anyone else. This irrational and false belief might motivate her when she builds houses, so that she can build them in fact faster just because she has this belief (even though some people are clearly faster than her, and on reflection she would come to realize this). Furthermore, it is conceivable that our eremite is quite rational when it comes to other topics, so that her epistemic character does not get spoiled by having one irrational belief.

It thus seems that Clifford's strict position is not defensible. I will call his position, which I from now on regard as being refuted, *Cliffordianism*. I choose this label because it will be essential to distinguish it from evidentialism proper.

Yet we should grant that having rational beliefs, that is, beliefs which are likely to be true, is *most times* good. True beliefs help us in guiding our lives according to the way the world is. Ignorance about the way the world is can get us into serious trouble. This, however, is not sufficient to support the strong claim that we should *always* have rational beliefs.⁶ Furthermore, rational belief is especially important in cases where the wellbeing of *others* depends on whether we get things right (Beckermann 2013, 33). But this is just a conceptual consequence of a quite trivial insight: In cases where it *matters* whether we have a *true* belief about

⁴ We must note here, however, that when a tendency towards irrationality can be useful, it is not clear why it should be bad to transmit a tendency towards irrationality towards other people (e.g., to your children by educating them in such a way that they are more self-confident than the evidence allows). Cf. also Schmidt 2017, 610.

⁵ The case of irrational belief about your partner shows that one's irrational beliefs might also be useful to someone else. However, cf. Crawford *fc* for skepticism about the utility of irrational beliefs in the context of relationships.

⁶ This is the main reason why accounts according to which evidence or epistemic reasons gain their normative force from the general practical value of true belief fail. For such a view, cf. Foley 1993. Another problem for such views is that we would have reasons to make our beliefs true if they are not yet true. However, this is implausible. For I would then, if I believe that I will die tomorrow, have a reason to ensure that I die tomorrow merely in virtue of the fact that I believe that I will die tomorrow (cf. Owens 2017b, 90). Kieseewetter (ms) calls this *the problem of the direction of fit* for accounts of epistemic reasons that ground their normative force in the practical value of complying with these reasons. For this problem, cf. also Raz 2011, 45.

something, we should do what we can do to ensure that we are epistemically rational. After all, being epistemically rational will make it most likely that we get things right.

As a general strategy, being epistemically rational is most effective if we want to make reasonable decisions based on true beliefs. Here it seems that what matters is not believing the truth *per se*, but rather believing the truth about those propositions where we need to know the truth in order to perform right actions. It seems that it is not *truth* which ought to have the highest priority in our doxastic lives. Rather, it seems that what we should strive to realize in believing is to have those beliefs that are *best* to have, all things considered, that is, those beliefs which will help us and others in making the right decisions. We should not always believe what is true, but rather believe what is best to believe, all things considered. For sometimes, we would be better off if we did not believe the truth.

Even if we grant this last point, of course it does not follow that truth is not the aim of believing. In order to properly understand the role of practical value and truth in our doxastic lives, we need to understand in what sense we *ought* to believe the truth, and in what sense we *ought* to promote valuable beliefs. Distinguishing between these senses will be one of the main tasks of the present book. In part II, I will argue that we need to distinguish two senses of ‘ought’ and ‘reason’ in order to make mental normativity intelligible.

However, contemporary evidentialists would part with Clifford, and the arguments against Clifford do not apply to their position. They argue that the *constitutive aim* of beliefs – an aim that is provided by the *concept* of belief – is truth.⁷ Even though we can evaluate beliefs as better or worse to have, evidentialists claim that evaluations of beliefs as prudentially or morally good are just irrelevant to what we ought to believe – *pace* Clifford. According to evidentialists, however, the fact that it would be, for example, pleasurable to harbor a certain belief will not provide us with a reason *to believe*, but rather with a reason *to bring this belief about* (if possible). For example, if Pascal tells us that we are better off if we would believe in God (Pascal 1670, §233), what we should do (assuming Pascal is right) according to evidentialists, is not believing in God ‘just like that’ for the reason that we are better off if we believe in God. For it seems that we cannot acquire a belief (directly) *for* such reasons. Furthermore, it is doubtful whether a belief acquired for such a reason (if that were possible) could still be seen as a subject’s *belief* (Williams 1970, 48-49).⁸ Rather, what we should do in a situation

⁷ Cf., e.g., Adler 2002; Kelly 2002; 2003; Wedgwood 2002; 2017.

⁸ Cf. chs. 8 and 9, where I argue that we cannot believe *for* state-given reasons. I will also point out that the possibility of causing a belief-state in ourselves is restricted by what other beliefs we have: A mental state cannot be seen as a belief of ours if it is in conflict with many of our other beliefs (cf. 8.1.3).

where we realize that having a certain belief would be very beneficial is to exercise *indirect control* over belief by taking certain measures in order to come have the belief: For example, if the belief in God is indeed as valuable as Pascal claims, then we should probably engage in religious practices, avoid talking to non-believers, and so on. If everything goes well, then soon the evidence we have might sufficiently support the belief that God exists. At this point, we ought to believe that God exists, given belief's constitutive aim. It is at this point that the belief will count as epistemically rational. And it is at this point where we can truly say, according to evidentialists, that we *ought* to have the belief. For now, the required evidence is available to us, and the evidence against god's existence is shut off from our perspective.⁹

1.2 THE ETHICS OF BELIEF, JAMESIANISM, AND INDIRECT DUTIES TO BELIEVE

If we consider the problems of Clifford's moral evidentialism, the position of Clifford's classical opponent, William James, does not seem to be as controversial anymore – or at least so I will argue now. James argues, as I did above, that Clifford was wrong in assuming that his principle holds when applied to *all* cases. According to James, when the evidence as to whether *p* is inconclusive and some further conditions are satisfied, we are, in a sense, allowed to let our 'passional nature' take control of our belief-forming processes, thereby ending up with epistemically irrational beliefs (James 1896, 11). Nevertheless, James follows Clifford when he accepts that truth is of central value for the formation of beliefs: when the evidence clearly speaks for or against a certain proposition, we never have the right to move into the contrary direction in forming our beliefs just because it would be more pleasant not to believe the truth or to believe a falsehood. Thus, James does not argue that we should always believe what is best for us to believe, and regardless of the belief's truth or likelihood to be true.¹⁰ James claim is more modest. It says that there are cases in which it is *allowed* to believe something that is not sufficiently supported by one's evidence.¹¹

It is essential to see that James does not say anything that would commit him to doxastic voluntarism, that is, to the claim that we can believe something just for the reason that it would

⁹ I will remain neutral as to whether epistemic rationality is a matter of responding to the evidence as it *seems* to you, or to the evidence which is – in a sense to be specified – *available* to you. One might wish to distinguish between several senses of more 'subjective' and more 'objective' rationality (Kolodny 2005, 510-511).

¹⁰ Meiland (1980) defends a stronger claim along these lines. I argue in 7.1 that there is a (trivial) sense in which these strong or radical versions of pragmatism are true. Yet I also show in chapter 7 how this radical position does not conflict with evidentialism if we understand it properly.

¹¹ For a version of pragmatism that understands itself as being more in line with the moderate Jamesian proposal cf. esp. McCormick (2015).

be better to believe it. Rather, his idea of *allowing the passions to influence our beliefs* suggests that we should adopt more indirect strategies in order to cause good beliefs in ourselves, like diverting our attention from certain facts to other facts, imagining the desired state of affairs vividly so as to make oneself believe in it, or cultivating an emotional life which allows our beliefs to be unresponsive to certain evidence in certain contexts.¹² Evidentialists would not have a problem with this. For they could just accept that it can be allowed, or even that we ought, to *cause* useful beliefs in ourselves, even though practical reasons are irrelevant when we ask *what to believe*. Such pragmatic permissions or obligations (understood as duties to bring valuable states of affairs about) are, according to evidentialists, restricted to *actions*, and not applicable to belief.

Here two points about Clifford and James come into focus which did not get the attention they deserve. Firstly, the only belief-control both authors had in mind is undisputed *indirect* control. When the debate about the ethics of belief was on the rise, the question about how we control our beliefs did not have the kind of sophistication it has today. Currently, philosophers discuss about different notions of *direct control* we could have over our beliefs, or exercise in believing, which are supposed to explain why we are responsible for our beliefs in a way we are not responsible for our headaches (which we can, arguably, influence indirectly). It seems to me that neither Clifford nor James ever sympathized with, or even considered for a moment, the idea we could control our beliefs more directly than through various bodily as well as mental actions.

Secondly, and related to the first point, both Clifford and James did not sharply distinguish between the questions of *what we ought to believe* and of *what beliefs we ought to bring about* (or, more generally, *how to manage our doxastic attitudes*). Clifford's essay clearly sets off discussing the first question, but nevertheless his first section is named, surprisingly, 'The duty of *inquiry*.' Inquiry, however, consists of activities which aim at forming true beliefs. Such activities amount merely to indirect control over belief. If we can fulfill our doxastic duties, that is, our duties about what to believe, by performing those activities which make up inquiry, then doxastic duties can be fulfilled by indirect control. Most contemporary ethicists

¹² Especially in section II of his essay, James points out the involuntary nature of rational belief-formation. However, it is not clear in how far he sees us to be in control when 'the passions' influence our beliefs, on which he elaborates on in section III. I do not see any textual evidence that he thinks that we have *direct* control in some cases where our passions influence what we believe. Rather, he says that in these cases 'we find ourselves believing, we hardly know how or why' (11).

of belief reject that doxastic duties can be explained by reference to indirect control.¹³ When contemporary philosophers talk about what we ought to believe, they mean that beliefs are directly subject to epistemic or practical norms, rather than that doxastic duties are, strictly speaking, just duties to bring certain beliefs about, and are thus reducible to ordinary duties to perform certain actions. Contemporary epistemology assumes that there are *genuine* doxastic norms, i.e., norms about what we ought to believe which are not just reducible to norms of doxastic managing-practices.

In contrast to contemporary ethics of belief, both Clifford and James were committed to the idea that at least some ‘doxastic duties’ are duties to bring a belief about, that is, duties to perform certain actions which result in beliefs. For the kind of belief-control Clifford and James had in mind, and which they deemed to be relevant to what we ought to believe, is *indirect control*. However, the topic they were concerned with, in their own words, was the question of *what we ought to believe*. It follows that they were committed to the idea that indirect control explains why there can be duties to believe. The ethics of belief is, understood in the way of Clifford and James, reducible to ordinary ethics.

It is surprising that both Clifford and James are still viewed as the traditional proponents of evidentialism and pragmatism, given that their topic were not genuine duties *to believe*, but rather duties to *manage* one’s doxastic life. The debate between contemporary evidentialism and pragmatism is, in contrast, a debate about *what to believe*, where this question is understood as independent of any indirect duties to manage one’s beliefs.¹⁴ This book will be mainly concerned with such *irreducible* norms of belief and *irreducible* norms of other attitudes, thereby contributing to the contemporary debate. Yet it is essential also for contemporary epistemology and metaethics to be aware of the distinction between indirect duties to manage one’s mind and norms applying directly to one’s attitudes (like the standards of rationality). It will avoid confusion by allowing us to understand the debate between evidentialism and pragmatism as a debate about the norms applying directly to beliefs.

¹³ Exceptions are Chrisman 2008; 2016; Meylan 2013; 2017; and Peels 2017. I think, however, that the connection between indirect control and doxastic duties they assume merely implies the uncontroversial form of pragmatism I present in 7.1. I think their accounts do not capture the essence of *epistemic* normativity (cf. ch. 4.4).

¹⁴ *Pace* some proponents of *The easy solution* (cf. ch. 0.1) mentioned in the last footnote. I discuss such views in chs. 2.2 and 4.4. Also some pragmatists would not agree with my diagnosis insofar as they disagree with a strict separation between the questions what to believe and what believe to bring about. Against this, I will argue that indirect control over beliefs is not sufficient to be able to believe for practical reasons in chapter 7. I argue that indirect control does not support any controversial form of pragmatism – it merely supports what I will call *non-traditional* pragmatism that does not conflict with evidentialism.

I will return to this distinction in chapter 3, where I specify how to understand the question of what attitudes we should have. For now, it is sufficient to have the distinction between norms that *directly* govern our attitudes and the requirements to *indirectly manage* our attitudes in front of us: the second type of norms are not norms that govern our attitudes, but norms that govern our actions by which we can influence and control our attitudes. If we talk about the latter norms, truth will not play a very significant role in defining the norms of belief. For the indirect norms of belief will be, in essence, just the requirements that govern any other actions: requirements of prudence and morality. Yet notions of truth and evidence might be essential for understanding the norms that directly govern our beliefs.

1.3 FROM THE ETHICS OF BELIEF TO THE ETHICS OF MIND

Questions about doxastic normativity and responsibility are discussed by contemporary scholars under the headings ‘Ethics of Belief’ (Feldman 2000), ‘Reasons for Belief’ (Reisner / Steglich-Petersen 2011), ‘The Aim of Belief’ (Chan 2013), or ‘Epistemic Norms’ (Littlejohn / Turri 2013). I think the key issues can be illustrated by the following sets of questions:

- (1) Is truth the only relevant value when we ask ourselves what to believe? Can the value of truth be weighed against practical values like the pleasure a belief might give us?¹⁵
- (2) Is believing for practical reasons, i.e. believing for reasons which indicate the goodness, value, or utility of having a belief, even *possible*? Believing for practical reasons seems to require that believing is itself action-like. Yet, though we can often do whatever we want, we cannot just believe whatever we want.¹⁶
- (3) But if beliefs are not like actions, how can we control our beliefs? Do we have any meaningful influence on or control over what we believe (that is, beyond indirect control)?¹⁷

¹⁵ I omit a long list of references on the first question mentioned under (1) here – this chapter contains references and I take up the issue again in chapter 3. The questions will be discussed at length throughout ch. 4 and throughout part III of this book. Reisner (2008) proposes an account how to weigh epistemic against pragmatic reasons for belief.

¹⁶ Ginet (2001), McCormick (2015), Rinard (2015; 2017), Reisner (2009) think that it is possible to believe for practical reasons; Bennett (1990), Hieronymi (2006), Scott-Kakures (1994), and Williams (1970) think that it is not. Evidentialists usually deny this possibility, too. I spelled out an argument against believing at will in Schmidt 2016 (German article). A significantly altered and extended version of this argument is found throughout part III.

¹⁷ Meylan (2013; 2017) and Peels (2017) argue that we can explain our responsibility for belief just by reference to our ability to influence them indirectly by means of our actions. McHugh (2013a), Montmarquet (1993), Owens (2000; 2017a), Hieronymi (2006; 2008) as well as Smith (2005; 2015a; 2015b) and others deny this.

- (4) If we lack any meaningful control over our beliefs, how can we justify our practice of criticizing others for their beliefs? How can we explain our responsibility for believing? Aren't we subject to the norms of epistemic rationality? If we fail to obey them, we end up being epistemically irrational. Isn't the charge of irrationality a form of criticism which presupposes control?¹⁸

The debate on belief normativity sketched by these questions is currently flourishing, as the (necessarily incomplete) references in the list show. At the same time, more and more authors start to look *beyond* the topic of belief and realize that almost the very same questions must be asked with respect to other mental states as well (see especially Schmidt/Ernst *fc*). Maybe we are not only responsible for believing, but also for *willing* (*desiring* and *intending*) and *feeling* (for our *emotions*). As various authors argued, we can neither choose our emotions at will (Oakley 1992, 130-132), nor our intentions (Kavka 1983; Owens 2000, 81-82; Hieronymi 2006). Nevertheless, we sometimes praise and blame people for what they feel and intend. We say, for example, 'You should no longer be angry with him because of what he did to you years ago' or 'I am proud of you that you intend to take your life in your own hands.' We show ordinary *reactive attitudes*, like praise and blame, resentment, indignation, or (self-)esteem, towards the attitudes of others and towards our own, which indicate our responsibility or even constitute it.¹⁹

An *ethics of mind* starts with the observation that we are responsible for our attitudes, or that they are subject to norms, and deals with questions analogous to those of an ethics of belief. Our practice of praising and blaming people not only for what they believe, but also for what they desire, feel, and intend seems to presuppose responsibility in these cases. But if responsibility requires that we control that for which we are responsible, then it follows that we control our attitudes.²⁰ Some questions arise naturally: How could this control look like, if it is not the *voluntary* control that we exercise when we act, and how can this control explain

¹⁸ Owens (2000; 2017a, Introduction) and Adams (1985) argue that our mental responsibility does not always presuppose control (and I join them in this book). Others, like Hieronymi (2006; 2014; *ms*), McHugh (2013a; 2014a), McCormick (2015, esp. Part II), Raz (2011, esp. his ch. 5 and Part III), and Smith (2005; 2015a; 2015b) think that it does.

¹⁹ As pointed out by Strawson (1962). Cf. ch. 0.2.3 on responsibility.

²⁰ The problem can be formulated with respect to many 'involuntary' mental phenomena in general (Adams 1985; Smith 2005). We seem to be responsible for involuntary failures like our attention and patterns of awareness, our emotions (cf. Neu 2010; Oakley 1992; Döring *fc*), for the thoughts that cross our mind in certain contexts (cf. Tognazzini *fc*; for some disagreement, cf. McCormick *fc*), and maybe even for our moods (cf. McCormick/Schleifer 2006a; 2006b).

our mental responsibility and the presence of norms that govern our attitudes?²¹ How does the kind of control we exercise in believing, desiring, feeling, or intending affect what attitudes we ought to have or what attitudes it would be rational to have? Should we always have those attitudes that are rational to have? Maybe it is sometimes better to be irrational? Could this mean that we sometimes *ought* to be irrational?²² These questions are strikingly similar to the questions mentioned above with respect to the ethics of belief. I will set out in the next chapter to characterize the two main questions of the ethics of mind in more detail and work out their relevance.

1.4 SUMMARY

This chapter introduced the ethics of mind by providing some historical background with a discussion of the positions of Clifford and James. We saw that Clifford and James understood the question of what we ought to believe in a specific way – as asking for our duties of belief-management. In contemporary epistemology, what is at issue are norms that directly govern our belief, rather than our belief-managing practices. This distinction will be helpful in getting a grip on the second main question of an ethics of mind in chapter 3. Finally, we saw that nothing prevents us to ask the questions that occupied the ethicists of belief also with respect to other attitudes. Doing so might gain new insights not only into the normativity of these other attitudes. Rather, the hope for epistemologists could be that thinking about the normativity of desire, emotion, intention, hope, phantasy, blame, decision, and other intentional mental phenomena might provide fresh ideas for understanding the normativity of belief as well. Of

²¹ Fischer and Ravizza (1998, 255-59) have argued that their account of ‘guidance control’ which we have over our actions can also be applied to emotions (for discussions on the normativity of emotions, see Döring 2010; fc; Neu 2009; Oakley 1992, chs. 4 and 5;). A defense of an approach explaining mental responsibility that is in line with this account can be found in Fischer/Tognazzini (2009). McCormick’s account for controlling beliefs (2015, ch. 6) is based on the theory of Fischer and Ravizza (I discuss her account in chapter 9). Other authors develop concepts of control which are intended as explanations of our mental responsibility in general, where these concepts are different from ordinary voluntary control we exercise when we act for reasons (Hieronymi 2006; Owens 2000 (his account of controlling intentions and decisions); 2017a, Introduction; Smith 2005; 2015a; 2015b). For an interesting discussion of the nature of intentions and the control we have over them, as well as on the connection between intention and action, see Shah 2008. On the kind of agency we exercise in deciding, cf. Soteriou fc.

²² The normativity of rationality is discussed extensively by Kiesewetter 2017; Lord 2018; Wedgwood 2017. An influential critique of the idea that rationality is normative is the paper of Kolodny 2005. I argue (in Schmidt 2017) that it is sometimes better to be irrational, and that there is thus a sense in which we sometimes should promote irrationality, and even a sense in which it can be said that we ought to *be* irrational. I return to this sense in ch. 7.1, where I point out how it merely supports an uncontroversial version of pragmatism, and is thus not in conflict with the recent defenses of the normativity of rationality, which commit to an evidentialist account of reasons for belief (although cf. Lord fc for some restrictions on his evidentialism).

course, this book cannot cover all these attitudes, but will focus mostly on belief, intention, emotion, and desire (the first three more than the fourth). Fully understanding the normativity of attitudes is a larger project (cf. Schmidt/Ernst fc).

Chapter 2

Mental Responsibility and Reasons for Attitudes

In this chapter, I spell out the first main question of an ethics of mind: How can we be responsible for our attitudes? I argue that the problem of mental responsibility (i.e., of responsibility for attitudes) is ultimately a problem about the status of reasons and rationality. I do this by first considering how we intuitively seem to lack control over our attitudes (2.1), and how we are intuitively responsible for our attitudes (2.2), before discussing how the absence of control is in conflict with our mental responsibility. I argue that, although we intuitively lack direct voluntary control over our attitudes (2.3), the lack of direct voluntary control over attitudes is not the core of the problem (2.4). Rather, the core of the problem consists in understanding how we can be responsible to the norms of rationality (2.5). This is intuitively not explainable by any form of control – whether the relevant form of control is taken to be voluntary or non-voluntary, direct or indirect. Finally, I present the possible solutions we might endorse in reply to the problem (2.6).

Thus, thinking about the first main question of an ethics of mind naturally leads us considering reasons and rationality, and thus to asking the second one: what attitudes *should* we have? The next chapter then turns towards how we need to understand this question in the context of the present inquiry, and motivates my own approach in resolving the problem of mental responsibility.

McCormick helpfully characterizes the conflict between responsibility for belief and the absence of doxastic control as follows:

Attributions of responsibility and other deontological judgments in the doxastic realm are puzzling. For much of what we believe is beyond our control; we cannot decide to believe the way we can decide to act. It seems that such lack of control should excuse us from responsibility and judgment (McCormick 2015, 77).

As McCormick notes, the problem is not restricted to *responsibility* for beliefs. It seems that *deontological attribution* with respect to belief is puzzling in the face of our lack of doxastic control. After all, in some sense it seems to be true that ‘Ought implies Can.’ But how can sentences like ‘You should not believe this nonsense’ or ‘There is no reason at all to believe this’ be true if there is no doxastic control, that is, no ‘Can’? The very possibility of an ethics

of belief, as well as the project of normative or deontological epistemology, seems to become questionable.¹

Are we dealing with one problem here, or with two? Is there a problem about responsibility on the one hand, *and* a problem about reasons, justification, and rationality, on the other? In this chapter, I will argue that the problem of mental responsibility is best understood as the problem of understanding the status of reasons for attitudes.

2.1 HOW WE SEEM TO LACK MENTAL CONTROL

Let us consider, first, why it seems that we cannot *control* our beliefs. Most of our beliefs seem to be passively caused by our environment. We experience the acquisition of a perceptual belief – like the belief that you are reading this text right now – as something that *happens* to us, rather than something we decide for. You do not stop and reflect upon whether you should acquire this belief after the text is in front of your eyes. Rather, belief comes immediately with perception. The same holds in cases where our perception provides us with ambiguous evidence. In these cases, we *refrain* from judgment quite automatically. To take a classical example:² If a tower seems to be round from distance and I know that I am not close enough for judging the tower's shape, I cannot just decide to believe whatever strikes me as pleasing to believe. Rather, I refrain from judging quite automatically. If, however, I am close enough to the tower, I will come to believe what shape it really is without any further contribution from me (except for walking towards it).

Even in cases where we reflect about what to believe, there seems to be no place for genuine *doxastic* freedom. Imagine a scientist, Lara, who wants to find out whether a certain substance is water. Lara conducts some experiments, thinks about them, and comes to believe that it is in fact water. It seems that the only activities Lara *did* here were the experiments and her intentional deliberating or reasoning about what is true. Lara did not perform *another* action when she was done with her thinking. She did not have to actively *form a belief*. Rather, her thinking *concluded* in the formation of a belief, but this concluding is not an additional activity she had to perform.

This is confirmed by the following observation. Lara could engage in the experiments she conducts and in her activity of thinking for *practical reasons*: She could decide to engage

¹ On how 'the deontological conception of epistemic justification' might become questionable, cf. Alston (1988). Owens (2000) views epistemic normativity to be threatened by the absence of doxastic control. More recently, McHugh (2014a) has presented the problem as a problem about the very existence of reasons for attitudes.

² Cf. Sextus Empiricus, *Outlines of Pyrrhonism*, I.32.

in them or refrain from doing them depending on whether it was *important* or *of interest* to know the nature of the substance. By contrast, she could not decide to conclude her investigation for such reasons: considerations about whether it would be good, valuable, important, interesting, or useful to believe that the substance in question is water could not guide the formation of her belief when she is done with her previous activities. Rather, she will form the belief based on her evidence that was uncovered by the preceding investigation.³

We can illustrate the lack of direct control over beliefs as follows. Imagine we have a manual telling us what to do in order to acquire justified, true beliefs about a certain topic. Let us assume that just *reading* this manual is not enough to justify our belief. It might be, for example, that the author does not provide the information that would justify the belief, because it is important for her that her readers come to believe because of their own effort, rather than just because she tells them that this is so and so. These manuals would mention various activities we should perform, like reading certain books, investigating into certain topics, conducting experiments, or thinking about certain issues. These manuals would not tell us, however, what to do *after* we are done with all those activities. They would not tell us *how to form a belief* after we are finished with thinking, experimenting, etc. This is because there is nothing left for us to *do* after we tried to find out the truth by performing those activities. Any further advice beyond the activities we had to perform will not make any sense. We just would not know how this advice should look like or what to do with it: ‘First, read a book about birds of the kind xy, then watch bird x and think about whether it has property E. Finally, form a belief about whether bird x has property E.’ If the last piece of advice is not meant as urging us to think even more about the issue until we come to believe something, it just would not make any sense as an advice.

Thus, there is a dilemma for the proponent of doxastic control: we either form our beliefs spontaneously – i.e., without any previous deliberation, as in the case of most perceptual beliefs; or we form them reflectively – i.e., with previous deliberation, as when Lara considers the evidence for whether the substance is water. In the first case, it seems that we do not

³ In line with this, Owens (2003) argues that our aims in forming beliefs cannot interact with our other aims we have as agents. Importantly, the influential critique of Steglich-Petersen (2009) of Owens’ argument is not in conflict with what I say here, for Steglich-Petersen claims that certain activities, which conceptually aim at forming a true belief (like inquiry or reasoning), can interact with our wider aims. A belief-formation in the sense I use this term here cannot do so, as Steglich-Petersen accepts. To accept that a belief-formation (in the non-intentional sense of the term) can interact with our wider aims is to commit to *pragmatism* – a position I cannot discuss in detail here, but which is one of the options in response to the problem of mental responsibility, as I argue below. Steglich-Petersen now seems to sympathize with pragmatism (cf. the contribution of Skipper/Steglich-Petersen in this volume).

exercise control when we form beliefs. When we form them reflectively, only our reflection is active, but not the formation of our belief that results from it. As Richard Moran points out: ‘[T]here is no further thing the person *does* in order to acquire the relevant belief once his reasoning has led him to it’ (2001, 119).⁴ In the case of spontaneous belief-formation, it seems that we withhold our judgment from what is true quite automatically. Our rational capacities just work the way they do – we do not exercise direct control over the results of their proper functioning. It seems that it is not *us*, but rather our evidential situation, together with the way our rational capacities function, that determines what we believe in each situation – in which we might happen to be either with or without our contribution.⁵

This is a dilemma not only for the proponent of direct *doxastic* control, but for the proponent of direct control *over attitudes* in general. Most of our attitude-forming takes place without us being reflectively aware of it. We spontaneously form not only numerous *beliefs* about our environment, but also other attitudes. We form desires and intentions to make it in time to the meeting, hopes that we will still make it, fears that we won’t, regrets that we did not get up earlier to make it in time, or feelings of anger and resentment directed at the person in front of us who did not continue to drive when they had green lights. In these cases, we would not even have the time to decide for our attitudes, even if we could. And even if we *have* the time to form our desires, intentions, and emotions reflectively, then we might again intentionally engage in the activity of thinking about what is good, right, fearsome, regrettable, or deserving of our anger and blame, but the attitudes that arise from such voluntarily controlled thinking are nothing we choose. Even if we actively think about how to decide, it seems that our resulting intention or decision is nothing that we could ever directly decide for.⁶

The proponent of direct doxastic control might object that, until now, I only considered cases of *rational* belief formation. Sometimes, our desires, passions, emotions, and moods influence what we believe in a situation. In the words of James, our ‘willing nature,’ that is, ‘all such factors of belief as fear and hope, prejudice and passion, imitation and partisanship, the circumpressure of our caste and set’ (James 1896, 9), influence, and often determine, what we in fact believe. Will this not make room for genuine *doxastic* freedom? It does not seem so, for, as James himself notes, even in these cases ‘we find ourselves believing, we hardly know

⁴ Here, however, Moran does not want to deny that there is direct doxastic control. Rather, he only wants to point out that this control cannot be understood in terms of performing an ordinary action called ‘belief formation.’

⁵ Cf. esp. Strawson (2003).

⁶ Cf. esp. Kavka (1983). On how we exercise agency in deciding, cf. ch. 9.3 and, in more detail, Soteriou *fc*.

how or why' (ibid.).⁷ Again, we might take certain measures in order to ensure that our passions have this or that influence on what we believe. But this is just indirect control over belief, not *direct* doxastic control. Manuals which would tell us how to deceive ourselves into believing *that p* by instrumentalizing our 'willing nature' might, for example, advise us to deceive ourselves by attending to this and that, and by avoiding attending to other things. But they would not advise us, after we are finished with those activities, to *form the belief that p*.

2.2 HOW WE SEEM TO BE DIRECTLY RESPONSIBLE FOR OUR ATTITUDES

What, one may wonder here, is it all about this *direct* doxastic control? Why is the fact that we have indirect control over our beliefs and other mental states through actions not satisfying? We must answer this question in order to get to the core of the problem, and to see which *kind* of responsibility we are seeking to understand by invoking control. For obviously, the problem is not that we do not control our attitudes at all: we are sometimes responsible for them in the way we are responsible for other consequences of our actions. After all, nobody denies that we can influence our mind indirectly.⁸

The main reason why the fact that we have indirect control over our attitudes seems only capable to provide an inexhaustive explanation of why we are responsible for our attitudes is that attitudes are *held for reasons*. This distinguishes them from non-intentional mental phenomena, like sensations. The question 'Why do you feel pain?' can only be answered by giving a causal explanation ('I fell from the roof'). It cannot be answered by giving *justifying reasons* why it would be *appropriate* or *rational* or *make sense* to feel pain now. For pain, understood as a mere sensation which can occur at various bodily parts, cannot be rational or irrational. By contrast, 'Why are you angry?' cannot only be answered by a mere causal explanation ('I talked to my neighbor, and this talk brought about certain brain states of mine which caused me to be angry'), but also by giving *justifying reasons* ('My neighbor said a lot of offensive stuff'). It can be more or less rational to be angry in a situation, and one's anger can thus be more or less understandable. By contrast, 'I understand your pain' will only make sense if 'pain' refers to a complex emotion, rather than to a mere sensation – there is nothing to understand about a hurt

⁷ It is interesting to note that James uses this to strengthen his point about the influence of our 'willing nature' on belief. Thus, his point about the arational influences on our beliefs is not meant to support the idea that we have voluntary control over our beliefs.

⁸ Levy (2007) is sometimes mentioned as an exception (cf. McCormick 2015, 84-86). However, even Levy grants that we can sometimes influence our mind indirectly in a controlled manner, which includes that we can foresee some of the consequences our actions have with respect to our mind.

foot (besides its causes and physiological processes in the body). In contrast to beliefs and other attitudes, sensations are ‘brute’ – they are not part of the space of reasons.

One might restate this observation by saying that attitudes are *constitutively normative*, while mere sensations are not. Attitudes are governed by certain *norms*, and if we violate them, we are irrational. There are norms constitutive of the concept of belief, desire, intention, and the various emotions. But there are no norms that are constitutive of the concept of pain, which would tell us when and why it would be rational to experience pain.

Saying that the norms of rationality are *constitutive standards* for an attitude means that the more irrational someone is, the less we can be sure what attitudes the person holds. If someone says that they believe that drinking water from the sink at midnight on a certain date has healing effects, but also believes in the epistemic authority of contemporary physics and thus accepts a naturalistic world-view, then we normally do not only regard the person to be irrational, but we might also come to doubt whether the person has one of the beliefs we ascribe to them. The more inconsistent someone’s set of beliefs is, the more doubtful it is *what* the person believes. More generally: the more a person violates the norms of rationality, the more doubtful it is what attitudes the person holds.⁹

The crucial point is that we have *indirect control* over our sensations even though our sensations are not rationally evaluable. We can influence whether we have pain or how much pain we will have by seeking or avoiding dangerous situations, or by going or not going to the doctor. We can cause ourselves to feel pain by tweaking ourselves. This kind of control makes us indirectly responsible for our sensations. But this is obviously not sufficient for pain to be subject to any constitutive norms. If indirect control is not sufficient for pain to be rational or irrational, why should indirect control explain that *intentional* mental states – i.e., attitudes that are directed at features of the world – can be rational or irrational? Viewing sensations and attitudes on a par with respect to control, that is, thinking that we only control them indirectly through actions, seems to be incapable of doing justice to the idea that attitudes are rationally evaluable. However, it seems that we are more *directly* related to our own attitudes as agents:

⁹ Saying that certain norms are constitutive for attitudes also means that attitudes can be individuated by reference to the norms by which they are governed. Anscombe (1957, sect. 2) provides such an individuation with her ‘direction of fit’-distinction between beliefs and desires: ‘beliefs ought to fit the world,’ but this is not true for desires. Rather, we ‘ought to fit the world to our desires.’ The Ought in question here is the Ought of rationality – if one realizes that it is false that p, then it is not rational to, instead of giving up the belief, continue believing it and to rather adjust the world to one’s belief. Also the debate about belief’s ‘aim of truth,’ (Chan 2013) – an aim that must be different from the aim of, for example, intending, but also different from the aim of guessing or suspecting, which arguably also aim at coming to a true guess or a true suspicion in some sense (Owens 2003, 290) – is concerned with the constitutive norms of belief.

we are capable of being irrational *merely by* believing, feeling, desiring, and intending, and it seems that we can be held responsible to the norms of rationality. As Angela Smith puts it: ‘[W]e are not merely producers of our attitudes, or even guardians over them; we are, first and foremost, inhabitants of them’ (Smith 2005, 251). That is, our agentic relation towards our intentional mental states, and consequently our responsibility for them, seems to be much more *intimate* than mere indirect control over them seems to allow for.

The problem here is, however, that it is not clear how *direct* control over attitudes is supposed to look like. Different conceptions of supposedly direct mental control have been brought forward.¹⁰ *Prima facie*, however, attitudes are not actions, and actions seem to be our only paradigm of what we control directly. The control we have over our actions is commonly called ‘voluntary control,’ that is, control by the will, or by our intentions or decisions. Speaking more precisely, our actions are *exercises of voluntary control*.¹¹ If this kind of control is the only control that can legitimately be called *direct*, then there seems to be no direct control over mental states – for we seem to lack voluntary control over the mind.

2.3 HOW WE SEEM TO LACK DIRECT VOLUNTARY CONTROL OVER ATTITUDES

The *prima facie* lack of voluntary, and thus direct control, over attitudes can be illustrated by what Matthew Chrisman has called the ‘no rewards principle’ (Chrisman 2008, 346): In a situation where you are offered a reward for believing something for which you have no, or contrary, evidence, you cannot believe it ‘just like that.’ Suppose someone offers you a reward for

¹⁰ I will argue in part III that our mental states are no exercises of *voluntary* control – as our actions are. I will not discuss the accounts of other forms of mental control. Rather, the hybrid solution to the problem of responsibility for attitudes (see chs. 0.1, 2.6 below) will undermine the main motivation to spell out any other form of mental control. For in one sense of ‘responsible,’ we are responsible for our mental states in virtue of indirect control; in another sense of ‘responsible,’ we are responsible for our attitudes in virtue of the fact that they are responsive to reasons. It is unclear why we should conceive of responsivity to reasons as implying control – at least in a sense where saying ‘we control our attitudes’ adds anything to the fact that they are responsive to reasons.

¹¹ Most authors prefer to say that our actions are *under* our voluntary control, or that we exercise voluntary control *over* our actions. This is misleading, however. It suggests that there are ‘exercises of voluntary control’ which are different from our actions themselves, and that we control our actions by means of these further exercises of voluntary control. Though I will grant that there might be a sense in which we exercise control when we intend or decide, intention or decision are not itself exercises of *voluntary* control – rather, they could at most be exercises of what Hieronymi (2006; 2014; ms) calls ‘evaluative control.’ The whole point with the use of ‘voluntary control’ is to highlight a kind of freedom which we enjoy only when we act. This is not to say that intentions and decisions are on a par with things that merely happen to us. It is merely to say that they are no actions: we do not exercise our agency in the same way when we act as we do when we believe, desire, feel, intend, or even decide something.

¹¹ However, Hyman (2015) recently argued that even inanimate things, like the sun, can act (it is *shining*). I reserve ‘action’ or ‘to act’ here for full-blooded agency that is intentional under a description.

believing that the number of stars in the galaxy is even. Arguably, not much hinges on whether you believe something about this issue, except that, in the current situation, you will get a lot of money if you acquire the belief. So why not just believe it in order to get the money? It seems to be – either psychologically or conceptually – *impossible* for you to acquire the belief *just like that*. You might deceive yourself into believing that you have evidence for the desired belief, or you might swallow a futuristic pill that induces beliefs, or you might try to convince the people offering you the money that you have the relevant belief, and act as if you believe it (e.g., you might reply with ‘yes’ when you are asked whether the number of stars is even, even though you do not believe it). However, in the latter case you do not really believe it, and in the former cases, you bring yourself to believe it by more indirect strategies.¹² It is hard to see how we could just directly form the belief that the number of stars is even. The reason for this will be uncovered in part III.

Some conclude from this that we *never* have voluntary control over our beliefs, that we can never ‘believe at will.’ They understand voluntary control over belief in such a way that having voluntary control over belief would require that we can believe actively *for* reasons of the kind provided by the reward in situations like these.¹³ However, if we do not have such voluntary control over belief, how can beliefs be subject to norms? After all, we (often) *can* perform an action just because we see that the action would be good to perform. When someone offers me a reward for lifting my arm, it is quite easy for me to do so and collect the reward. Arguably, the fact that we can control actions in this way (‘voluntarily’) is what explains (at least partly) why we sometimes *ought* to do one thing rather than another, why there are *reasons* for actions, and why we can be *blamed* and *praised* for what we do.

Forming an emotion or an intention just because it would be good to have the emotion or intention seems at least as problematic as forming a belief for such reasons.¹⁴ This *prima facie* problem can be experienced by each of us if we imagine situations analogous to the one described above: If someone offers you a reward for desiring something completely undesirable (either something completely neutral or something very bad), or for merely *intending* an

¹² I think that self-deception (as an *activity* on some level) can be conceived as changing your evidential situation in some sense: you attend to certain evidence and ignore other evidence. The role of attention-control in self-deception, and implied by this, the active aspect of deceiving ourselves, is highlighted by Fingarette (1969).

¹³ I will refer to the relevant reasons as *practical reasons*, which are a subtype of state-given reasons. For this terminology, cf. ch. 0.2.4.

¹⁴ For the problem with respect to emotions, see Oakley 1992, ch. 4. For the problem with respect to intentions, cf. Hieronymi 2006, Kavka 1983, Owens 2000, 81-2, and Pink 2009, sect. 6.

action for which you have no reason to perform it,¹⁵ or for being angry about someone who did not do anything bad to anyone, or for fearing something which you do not consider as dangerous or fearsome, then you cannot desire, intend, or feel ‘just like that.’ I will not go into detailed descriptions of examples for every attitude, for the examples are easy to imagine, and they have been described in detail by others (see footnote 14 above). It is only important at this point to note that, *prima facie*, it does not seem to be the case that we have direct voluntary control over mental states. Thus, the question arises of how mental states can *ever* be subject to norms.

2.4 HOW THE LACK OF VOLUNTARY CONTROL IS NOT THE CORE OF THE PROBLEM

How could the *prima facie* lack of control over attitudes pose a problem for our practice of holding each other responsible for our attitudes? It is helpful to state the problem as a conflict of three claims which, when all true, would result in a contradiction. Thus, to resolve the conflict, we have to reject at least one of the three claims.¹⁶ A first try would be to state the problem as follows: (1*) We are responsible for our attitudes; (2*) We are responsible for our attitudes only if we can control them; (3*) We cannot control our attitudes.

The problem with this way of formulating the puzzle is that (3*) would be obviously false. As mentioned on various points before, we have at least *indirect* control over our mental states: we can investigate and engage in thought and reasoning. We can also control our mind indirectly by engaging in projects of acquiring, say, true beliefs and other rational attitudes about Sankt Petersburg by going there and walking through the city. So, this first attempt of formulating the problem does not get at its core.

Given that it is not *any* control that seems lacking, but *voluntary* control, that is, the kind of *direct* control we have over our actions, we could put (3*) not in terms of control, but rather in terms of *voluntary* control. Still, this would not get to the core of the problem. Note that our exercises of voluntary control – our actions – are subject to norms of prudence and morality. If we had voluntary control over beliefs, we could evaluate someone’s beliefs in terms of how beneficial they are to the person who has them, or in terms of how harmful they are to others. Yet we would not be justified in holding people accountable for their *epistemic* failures: we would not be justified in regarding someone as criticizable or blameworthy for their

¹⁵ Assume that you know that you will not get the reward if you will actually perform the action: you receive the reward only if you *merely* intend the action without actually doing it. Without some roundabout routes (like ensuring that you will not perform the action in the future and making yourself forget about this fact), you will not be able to intend the action insofar as you are rational. The point is similar to Kavka’s (1983).

¹⁶ I owe the idea of stating philosophical problems in such a way to Gerhard Ernst, cf esp. Ernst (2008, 65-71).

epistemic irrationality. For we would only be justified to blame or criticize someone for their irrationality if this irrationality is bad either in terms of prudence or morality. Yet it seems that we are sometimes justified to blame or criticize people *for failing to properly base their beliefs on their evidence* rather than for failing to have beliefs that are not harmful for themselves or for others.

One reply is, of course, to just deny that there is such a thing as epistemic blame: we never blame people merely for being epistemically irrational; rather, we blame them for irrationality only if their irrationality is harmful to themselves or to others. I will return to this option of denying the existence of epistemic blame as well as other forms of blame for holding in part II of this book. For now, it is sufficient to note that denying the criticizability of irrationality would be highly controversial in the context of contemporary philosophical debates. The claim that irrationality is criticizable is often viewed as a starting intuition for defenses of the normativity of rationality (cf. Kiesewetter 2017, chapter 2; Lord 2018, 3-4). If we are criticizable or blameworthy (in what follows, I will stick with the label ‘blame’ instead of ‘criticism’ for the sake of brevity) for such failures in our rationality, then voluntary control over attitudes would provide an extensionally inadequate explanation of our blameworthiness for irrationality: we would be blameworthy only if our irrationality is prudentially or morally bad.

That the norms of rationality – as I understand them, and as (I take it) most of us understand them intuitively – cannot be made intelligible by reference to voluntary control can also be illustrated without reference to the notion of blameworthiness. For if we could believe or feel at will, that is, believe or feel whatever we want to, it would be the case that we *should* purposefully ignore the norms of rationality in order to achieve some good. Take the following example of a norm of rational belief to illustrate this: We ought to believe only what is sufficiently supported by our evidence.¹⁷ If, in a given instance, we would be better off believing something that is not supported by sufficient evidence, and we have voluntary control over what we believe, we could just ignore this norm in order to gain the desired good. Furthermore, it seems that, if we are *better off* if we believe it, and if we *can* just believe it by an exercise of voluntary control, we *should* do so.¹⁸ In this case, we should exercise voluntary control in a

¹⁷ Cf., for example, Owens 2017b, who provides an interesting proposal for how to spell out the ‘sufficiency’ required here. As I will discuss in ch. 4, the norm might require a more careful formulation. There are also reasons to doubt that there is such a purely epistemic norm governing our beliefs. However, I will ultimately defend the idea that we can be held accountable to such purely epistemic norms.

¹⁸ This claim needs at least one restriction in order to be true: The exercises of voluntary control we could perform must be *reasonable*. This includes, for example, that there are not more important things to do than forming the corresponding belief (e.g., visiting your father in the hospital). I will discuss the implications of the claim that we should sometimes act in such a way that we violate the norms of rationality in ch. 6.

way that violates belief's constitutive standards. Thus, it seems that our exercise of voluntary control is not governed by the norms of attitudinal rationality, but rather by practical norms (of prudence and morality) which also govern our actions.

Voluntary control over attitudes would thus not explain why attitudes are subject to the norms which are most characteristic of them. Rather, it would make the idea more plausible that our attitudes are instead directly subject to norms of prudence and morality, as our actions are. Insofar as the norms of rationality cannot be equated with the norms of prudence and morality which govern our actions – and this seems most obvious in the case of belief's constitutive norms, which are (or at least *entail*) purely epistemic norms –, the presence of voluntary control over attitudes would not explain why we are held responsible to these norms.

One might object that belief's constitutive norms are not epistemic, but that they are rather prudential and moral norms which together aim at forming those beliefs which are practically right to form. For otherwise, how can we ever morally blame people for objectionable beliefs? I do not exclude this possibility from the start. Later (chapter 11) I will even argue that we *can* blame people for their beliefs in some sense: we morally blame them if their beliefs are manifestations of disrespect.

For now, our aim is to formulate three *intuitively plausible* claims that lead into contradiction. The idea that beliefs are subject to moral norms (or even *requirements*) is *prima facie* not plausible given the very idea that epistemic standards are the constitutive norms of belief. It seems to be the specific nature of epistemic standards which differentiates beliefs from other attitudes. Hieronymi points out that in believing that *p*, I am committed to *p* as true, while I am not so committed when just supposing or imagining that *p* (2006, 49-50). When imagining that *p* or supposing that *p*, I am also imagining or supposing *that p is true*. But I am not thereby committed to an answer to the question of *whether p is true*. In believing, by contrast, I am committed to having answered this very question. Thus, it will be legitimate (in some sense) to request my reasons which bear on the question on whether *p* is true, simply in virtue of the fact that I believe that *p*. The norms which determine the rationality of believing, and in the light of which my answer to the legitimate request are evaluated, are thus epistemic standards that are constitutive of the concept of belief.

Neither can constitutive norms of other attitudes be explained by reference to direct voluntary control. Such norms prescribe (very roughly) believing the truth, desiring the

desirable, fearing the fearsome, admiring the admirable, and intending right actions.¹⁹ At the same time these constitutive norms are different from the norms which we should follow if we had voluntary control over the mental states. If we could just decide what to believe, desire, fear, admire, or intend, then we could and should sometimes use this power to intentionally flout the attitude's constitutive norms. Just think of a (sci-fi) case where you could prevent a lot of other people's suffering by voluntarily adopting irrational attitudes. In these cases, we should exercise our voluntary control so as to make ourselves irrational. Thus, it seems that voluntary control cannot explain why there are constitutive norms, including standards of rationality.

To get the problem started we would need three claims which are *prima facie* plausible but which at the same time lead into contradiction. However, as I have argued in this subchapter, the claim that direct voluntary control over attitudes would explain mental responsibility is confronted with two related problems. First, the claim provides a *prima facie* extensionally inadequate account of our blameworthiness. We seem to be blameworthy or criticizable as soon as we are irrational, and independently of whether our irrationality is harmful to others or ourselves. Secondly, direct voluntary control over attitudes would imply that our attitudes are subject to practical requirements. This is not intuitively plausible. If we are seeking for an explanation also of why we are responsible for whether we comply with the standards of rationality, then voluntary control, whether direct or indirect, is not suited for providing such an explanation.

We thus need a different formulation of the problem that does not invoke the notion of voluntary control. However, I will have to consider the idea that practical norms for belief-formation determine belief's rationality in more detail later (see part III).²⁰

¹⁹ These are (roughly) the norms of correctness, fulfilling which is the aim of rationality (cf. Wedgwood 2017). To be rational is to make it likely (under normal circumstances) that one reaches the aim of correctness (i.e., truth, desirable, fearsome, admirable, right action, etc.). Both the norms of correctness and the norms of rationality are constitutive norms insofar as they are inherent to the concept of the specific attitude. What I say in the paragraph above holds for both the norms of correctness and the norms of rationality.

²⁰ One way of arguing that voluntary control over attitudes is governed by constitutive norms of attitudes (like epistemic norms) rather than by practical norms would be to adopt some kind of *deontologism* with respect to the norms of attitudinal rationality: As it is true that killing someone might in some instances maximize value, this does not make it right or even allowed to kill someone; instead, killing is morally wrong in each and every case *because it is killing*. In the same way, it could be argued that believing what is epistemically rational to believe is what we ought to do *even if* we can decide what to believe: believing irrationally is always wrong *because it is believing irrationally*. However, I find this line of thought hardly compelling as soon as we accept that the consequences of what we do (or in this case: of our exercises of voluntary believing) *partly* determine what we ought to do (or believe).

2.5 THE CORE OF THE PROBLEM: RESPONSIBILITY TO THE STANDARDS OF RATIONALITY

Since the problem of mental responsibility calls for an explanation (at least *also*) of the kind of responsibility we have for (non-)compliance with the standards of rationality, and since it seems that this responsibility can be made intelligible neither by direct voluntary control (cf. 2.4) nor by indirect voluntary control (cf. 2.2), we should now try the following formulation: (1**) We are responsible for (non-)compliance with the standards of rationality; (2**) We can only be responsible for such (non-)compliance if we have *direct non-voluntary* control over our attitudes; (3**) We do not have such direct non-voluntary control.

I think this is a better way to state the problem. Yet we might wonder why (2**) is true as stated. For what is *non-voluntary control* supposed to be? Intuitively, the control we exercise when we act – direct voluntary control – is our only paradigm of direct control, and this is the kind of control that (at least partly) explains why we are responsible for our actions and their consequences. So how could a *non-voluntary* control explain responsibility? Since it seems that we do not know much about the nature of this control, we also do not see, *prima facie*, how it could do this trick. It thus seems that, intuitively, there is *no* form of control – direct or indirect, voluntary or non-voluntary – that could explain how we are directly responsible for being (ir)rational.

If we are indeed responsible for complying with these constitutive standards, i.e., the standards of rationality, and if it is not obvious what form of control could explain this, then we have an interesting version of the problem of mental responsibility:

- (1) We are responsible to the standards of rationality.
- (2) We can only be responsible to the standards of rationality if there is *some* form of control that could explain why we are responsible to the standards of rationality.
- (3) There is no form of control that could explain why we are responsible to the standards of rationality.

The plausibility of (1) derives from the intuition that being irrational – failing to respond correctly to one’s object-given reasons – is blameworthy in some sense. This implies that we are responsible for (non-)compliance with the standards of rationality: we are legitimately subject to various reactive attitudes in virtue of our being rational or irrational. As a shorthand formulation, I will say that we are responsible *to the standards of rationality*.²¹ The plausibility of (2) derives from the intuition that responsibility requires control, or ‘Ought implies Can’ – a

²¹ Cf. ch. 0.2.3 on this phrase.

claim that is, in one version or another, accepted also by philosophers concerned with the standards of rationality (cf. Kieseewetter 2017, 28; Wedgwood 2017, chapter 3). I have already motivated the idea that there seems to be no such thing as direct control over our attitudes (in 2.1 and 2.3), and that our familiar paradigm of control – voluntary control (direct or indirect) – cannot do the job of explaining why there is such a thing as responsibility to the standards of rationality (in sections 2.2 and 2.4). This motivates (3). In order to resolve our philosophical perturbation caused by the three claims, we need to find reasons for rejecting at least one of them.

Thus, all the premises are intuitively plausible. There are at least two other advantages of understanding the problem of mental responsibility in this way. First, we know better *what it is* that requires an explanation. Merely saying that we need to explain how we can be responsible for our attitudes leaves it unclear what exactly it is that requires an explanation. For it is obvious that we can be said to be sometimes responsible for our attitudes in the same way as we can be said to be sometimes responsible for other consequences of our actions. We are indirectly responsible for our attitudes in this way insofar as we have *indirect control* over our attitudes: we can intentionally reason to beliefs, manage our emotions by meditation, and determine our intentions and decisions by intentionally thinking about what to do. If we understand the problem as merely requiring an explanation of our *responsibility* for attitudes without specifying what *kind* of responsibility we have in mind, then it is unclear why indirect voluntary control does not provide a satisfying explanation. Yet, as I have argued above (in chapters 2.2 and 2.4), it is intuitively unclear how our capacity to indirectly control our mind can explain why we are responsible for (non-)compliance to the standards of rationality. Indirect responsibility for attitudes seems thus not to *exhaust* our responsibility for them. Since we do not see what kind of control could explain the special responsibility to the standards of rationality, we are faced with our problem.

Secondly, stating the problem in this way has a *metaphilosophical* advantage. For it shows us how different debates are relevant for solving the problem. For one, we need to think about the nature of responsibility and control. However, we also need to think about *the nature and the status of the standards of rationality*. Thinking about their status means to think about how we are responsible, and thus sometimes blameworthy, for (non-)compliance with our reasons for attitudes. Getting clear about this kind of responsibility might give us a clue as to what kind of control we are searching for. Thus, theorists of responsibility need to think about rationality and reasons, and theorists of rationality and reasons need to think about responsibility. As long as we do not understand in what sense we are directly responsible for complying with

our reasons for attitudes, and thus in what sense we can be legitimately blamed for irrationality, we do not understand the status of these reasons.

2.6 SOLUTIONS TO THE PROBLEM: HOW THERE COULD BE AN ETHICS OF MIND

We could reply to our problem by denying (1), i.e. rejecting the idea that we are responsible to the standards of rationality. The straightforward way to do this is to deny that there are *reasons* for attitudes at all. If the standards to which we are intuitively held responsible in fact do not exist, then we cannot be responsible to these norms. Call this response *mental nihilism*.²² Endorsing it would mean to fully commit oneself to *The easy solution*. But nihilism is a very radical response which is considered as ‘too drastic to be plausible’ (Wedgwood 2017, 72). I agree that we should not endorse such a revisionist view too hastily before having considered alternatives. Yet I will illustrate throughout this book (beginning with part II) how taking this radical response seriously can lead us towards a proper understanding of how we are held responsible to the standards of rationality.

A less radical way of denying (1) would be to argue that there are standards of rationality, but that we are not responsible to them. Call this response *anti-responsibilism*. The advantage of such a position is that it takes away the justification for the idea that being subject to the standards of rationality requires that one is in control of one’s attitudes. For this idea derives its plausibility from the intuition that responsibility requires control. If we are not responsible to the standards of rationality, then these norms could be placed on us independently on whether we have control over whether we comply with these standards. Such an account of our responsibility to the standards of rationality could understand rationality, for example, as ‘normal functioning of an agent.’²³ One way to defend such a position is by saying that rational criticism is comparable to criticizing defects in natural development, as when we ‘criticize’ an oak for not having deep sturdy roots (because that is how we expect oaks normally to be).²⁴ An agent that is irrational, according to such an account, would be an agent that is not how we

²² I adopt the label from Kiesewetter (ms). He distinguishes it from a view that denies the *normativity* of reasons for attitudes but accepts that they are still *reasons* – they are just not *normative* reasons, according to this view. This view is equivalent to *anti-responsibilism* I present below (if one thinks, as I do, that we are responsible to those and only those norms that provide us with normative reasons). I agree with Kiesewetter’s diagnosis that denying the normativity of these reasons is to deny their status as reasons (Kiesewetter argues that there is no plausible account of non-normative reasons for attitudes available). I return to the idea that reasons (in the sense I am concerned with them) necessarily have ‘normative force’ at length in part II, and I spell out how they have this force in part IV.

²³ For an account along these lines, cf. Raz 2005a; 2005b. For discussions and criticism, cf. Kolodny 2005, 551-556; Kiesewetter 2017, 33-36.

²⁴ I take the oak example from Broome (2013, 12), who took it from Foot (2001, 46).

normally expect agents to be. One problem of this position is that it seems to take the normativity out of rationality. For obviously, there is nothing substantially wrong with not being how others normally expect us to be.²⁵ The oak is not criticizable merely in virtue of the fact that it does not develop as it ‘ought,’ and neither would we. Anti-responsibilism is threatened to collapse into mental nihilism. To avoid nihilism, anti-responsibilists have to make it intelligible how there can be reasons for attitudes even though we are not responsible for complying with them. An oak is neither responsible for its defects nor does it have reason not to be defective.

I will argue that anti-responsibilists cannot make the act of balance between nihilism and responsibilism they have to make in order to uphold their position. In chapter 4.1.3, I will point out that a very weak conceptual connection between blameworthiness and reasons already makes anti-responsibilism collapse into mental nihilism. In chapter 11, I argue that being answerable for one’s attitudes is being responsible to the standards of rationality. The idea that our attitudes are in the space of reasons commits us to accepting that we are directly responsible for them (in the sense of being directly answerable for them). This is why anti-responsibilism fails.

We could also reply to our problem by denying (3). One way of denying (3) is to spell out a notion of control that *can* explain how we are responsible to the standards of rationality. Given my argument throughout this chapter, this kind of response seems to require us to spell out a notion of *direct non-voluntary* control which we exercise whenever we believe, desire, feel, or intend. As a response to how there can be responsibility for *belief*, this response is often called *doxastic compatibilism*, because it argues that the lack of voluntary control over belief is compatible with our responsibility for belief. Yet there are other responses which endorse this compatibility and which do not spell out a non-voluntary form of control (as I will explain below). However, for the sake of having a unified terminology within philosophical debates, and in the absence of any better label, I will refer to any position that explains our responsibility to rational standards through direct non-voluntary control as *mental compatibilism*. The label is also fitting insofar as mental compatibilists often seek for inspiration by considering traditional compatibilist accounts of free will. Philosophers defending mental compatibilism claim that we control our attitudes, e.g., by making judgments about reasons (Smith 2005), by

²⁵ For a more nuanced criticism, cf. Kiesewetter 2017, ch. 2.4. It is important that the notion of an ‘expectation’ here is purely epistemic. Just because I do not end up doing what others believe that I would probably do does not make me criticizable in any way. Matters would be more complicated if by ‘expectation’ we mean a more normative notion – a demand that is placed on us by others. But that is not how proponents of a ‘normal functioning approach’ would understand the expectation that someone be rational – otherwise normativity would be back in.

answering questions about what is true or important (Hieronymi 2006; 2008; 2014; ms), or by exercising guidance control (McHugh 2013a).²⁶

In this book, I will not discuss notions of direct non-voluntary control over attitudes.²⁷ This is because I think spelling out such notions is not helpful for resolving our philosophical puzzlement. The puzzlement arises because we do not understand how we can be responsible to the standards of rationality. Rather than assuming that this responsibility is to be explained in forms of control (i.e., rather than assuming (2)) as a working hypothesis of our investigation, it might be more fruitful to try to understand the status of the standards of rationality without making any assumptions about control from the beginning. The hope is that if we understand this status, then we can better evaluate what kind of control – if any – is presupposed by our responsibility to the standards of rationality. Here is a spoiler: it will turn out that we need not assume any form of control to understand this kind of responsibility. The mere fact that our attitudes are responsive to reasons – in a very weak sense of ‘responsive to reasons’ that is co-extensive with our being answerable for them – is sufficient for responsibility.

However, one could also object to my arguments which I used in support of (3). One could argue that indirect voluntary control *can* explain why we are legitimately held responsible to the standards of rationality. Maybe there are specific *epistemic actions* – like maintaining a coherent system of beliefs (Chrisman 2016) – that can explain epistemic responsibility. One might argue that we are not be able to exercise these actions over brute sensations, and thus these positions could acknowledge a distinction between reasons-responsive and non-reasons responsive states without committing to the idea that we have direct control over these states. The former states are responsive to reasons in virtue of the fact that we have a *special* form of indirect control over them. I will call any response that explains responsibility to the norms of rationality in terms of indirect control *indirect mental voluntarism*.²⁸

²⁶ Hieronymi (ms) and Boyle (2013) argue that our attitudes are *active states*. We are active *in* believing, desiring, feeling, and intending, as we are active *in* acting (though in another sense of ‘active’).

²⁷ My discussion of McCormick’s notion of ‘guidance control’ in ch. 9.4 might look like an exception of this. However, since I regard voluntary control as the kind of control that we exercise when we act, and since we exercise guidance control when we act, my notion of voluntary control encompasses guidance control. Yet I argue in 9.4 that there is a case to be made to distinguish between two forms of control that we exercise when we act.

²⁸ Meylan (2013; 2017), Chrisman (2008; 2016; ms) and Peels (2017) defend such an approach for explaining responsibility for belief. Chrisman is most explicit about the idea that his account is also intended as an account of why there are *epistemic norms* (rather than doxastic responsibility). Furthermore, it seems that Oakley (1992) could be read as indirect voluntarist with respect to emotions, for he is talking about responsibility for *emotions*, and not only for actions which cause emotions. Yet it is likely that Oakley would deny that we are always responsible for our attitudinal (ir)rationality.

I have already voiced my doubts about these positions, and I return to these doubts in chapter 4.4. Yet my argument in chapters 10 and 11 that we are directly responsible for our attitudes in a way that does not presuppose indirect voluntary control will also provide us decisive reasons for rejecting indirect mental voluntarism.

Or one might argue that *direct* voluntary control can, contrary to first impression, explain our responsibility to the standards of rationality. I call this response *the direct voluntarist's explanation*.²⁹ This reply would most likely involve an argument for *pragmatism*: that the constitutive norms of belief are not purely epistemic. Rather, we can sometimes believe for reasons that show the belief-state to be valuable – i.e., practical reasons. This is an additional burden this response is taking up: it has to argue, first, that our attitudes can be exercises of voluntary control, and, secondly, it has to show how the rationality of belief is (at least partly) determined by practical reasons.³⁰ These two tasks might turn out to be closely connected. However, on its own, the direct voluntarist's explanation will not be a satisfying response to our problem. This is because our attitudes are, intuitively, not only subject to practical norms. Beliefs, for example, are intuitively also subject to epistemic norms. Voluntarists thus also need an account of how we can be responsible to these epistemic norms. They can either deny the existence of epistemic norms altogether (Rinard 2015; 2017), thereby endorsing nihilism with respect to epistemic norms; or they could argue that these norms are such that they do not presuppose control, thereby endorsing anti-responsibilism; or they could endorse any account that rejects either (1) or (3) for epistemic norms.

However, I will argue throughout part III that we lack direct voluntary control over our attitudes, and that there are thus no practical reasons for attitudes. This is why also the direct voluntarist's explanation fails.

Last but not least, we might reject (2). I will call any response that denies that our responsibility to the standards of rationality requires control *mental involuntarism*.³¹ Rejecting (2) would require us to make this responsibility intelligible in a way that does not reduce the standards of rationality to mere norms of normal or natural development (thereby avoiding

²⁹ The label is a bit weird. I choose it because I will later (esp. in part III) be concerned with the claim that we exercise direct voluntary control over our beliefs, which I will call 'mental voluntarism.'

³⁰ I read McCormick (2015) as arguing in this way. For she argues also that the rationality of hope is governed by practical reasons (2017). I take McCormick's notion of 'guidance control' – a bit simplifying – to be a type of direct voluntary control. This is because I understand voluntary control as the kind of control we exercise when we act for reasons. And guidance control is meant to be exercised in acting for reasons as well. I will discuss the notions of voluntary control and guidance control in more detail in ch. 9.

³¹ Adams (1985) and Owens (2000; 2017a) are involuntarists (Owens especially with respect to beliefs, but not with respect to intentions). Both authors claim that we are *blameworthy* (in some sense) for not complying with the norms that govern our attitudes.

anti-responsibilism), and yet we need to understand the standards of rationality in such a way that they can be placed on us without requiring any form of control (thereby avoiding mental compatibilism, indirect voluntarism, or the voluntarist's explanation). Finding the right middle ground might seem like an impossible act of balance.

Yet my solution is a denial of (2) along these lines, and I thus claim that this act of balance is possible. As I pointed out in chapter 0.1, I will argue that we are never directly blameworthy for our attitudes in such a way that emotions like resentment, indignation, anger, or guilt are appropriate towards us *merely for having an attitude*. Remember that it was in this sense that I endorse *The easy solution*: in a sense of 'responsible' in which responsibility can be defined in terms of the reactive sentiments, we are responsible for our attitudes only in virtue of the fact that we have indirect voluntary control over them. There is no such *robust* sense of 'responsible' in which we are directly responsible for our mental states – i.e., no sense in which we can be fitting targets of passionate forms of blame *merely for having an attitude*. This is the first half of the act of balance: violating rational standards does not give rise to forms of blame that are typical for violating practical requirements.

The other half is that we *can* be directly responsible for our mental states *in another sense* of 'responsible.' We are justified, for example, to view our relationships with racists or sexists as impaired in certain ways merely in virtue of the fact that they have *disrespectful attitudes*, i.e., in virtue of the fact that they are irrationally disrespectful. Viewing relationships as impaired in this way is a form of (non-passionate) blame as well. Thus, there is a form of direct blameworthiness, and thus responsibility for attitudes. Yet understanding it properly will not imply that it presupposes control. This is the other half of the act of balance.

My solution thus has two main elements: *The easy solution* (under a certain reading of it) and mental involuntarism. We cannot view our relationship towards an oak as impaired when it is not an oak with deep sturdy roots, but we can sometimes view our relationships with fellow human beings as impaired if they fail to respond correctly to reasons for certain attitudes. That we can view our relationships as impaired in virtue of someone's irrationality is the sense in which reasons for attitudes have *normative force*, why it can be said that we *ought* to comply with them, even when we lack any control over whether we comply with them (mental involuntarism). However, this form of relationship-impairment does not yet warrant the reactive sentiments of resentment, indignation, anger, or guilt. This is because violating the standards of rationality does not have the same normative consequences as violating practical requirements. The reactive sentiments – resentment, etc. – are rational responses to our attitudes only if we had indirect voluntary control over our attitudes (*The easy solution* under a certain

reading). The mistake that is made by indirect mental voluntarism is to confuse this type of responsibility, which in fact presupposes indirect control, with our responsibility to the standards of rationality, which does not.

By endorsing mental involuntarism, I do not deny that it can be methodologically justified to spell out a notion of direct non-voluntary control over our attitudes that still warrants the label ‘control.’ My claim is merely that, in order to resolve our philosophical puzzlement created by the problem of mental responsibility, we do not need to assume such a notion of control. Acknowledging that our attitudes are (at least normally) within the space of reasons, and thus ‘reponsive to reasons’ in a very minimal sense that distinguishes them from brute sensations, is sufficient for understanding how we can be responsible for them. There can be philosophical argument about whether an attitude’s being responsive to reasons means that one is in control of the attitude,³² but this issue need not be decided in order for responsibility for attitudes to be intelligible to us. Instead of discussing extensively about forms of direct control, we will fare better, I propose, by trying to understand the nature of this responsibility and how it differs from the responsibility we have to practical requirements of prudence and morality. There might be other problems besides the problem of mental responsibility that require us to acknowledge that our attitudes are not completely passive.³³

In addition to the fact that my formulation of the problem is not only about *doxastic*, but rather about *mental* responsibility, my way of putting the problem differs in two further important ways from statements of the problem found in the contemporary philosophical literature. First, it is formulated in terms of *some form of control*, not just in terms of voluntary control. The reason for this is, as I have argued, that it is hard to see, *prima facie*, how our responsibility to the standards of rationality could be explained by any form of control, whether direct or indirect, voluntary or non-voluntary. A solution to the problem thus requires an explanation of how the form of control identified by the solution – direct or indirect, voluntary or

³² Owens (2000; 2017a, Introduction) denies this: according to his account of mental responsibility, beliefs and emotions are responsive to reasons but not under our direct control in any sense. Intentions are, according to Owens, responsive to reasons and under our ‘reflective control’ that is distinct from voluntary control.

³³ Hieronymi (ms) argues that a problem inherent to our concept of agency requires us to acknowledge that our attitudes are, in some sense, exercises of control. I do not reject this. I just reject that in order to explain how we are directly responsible for our attitudes, we do not have to assume that our attitudes are exercises of agency. It is sufficient to acknowledge their reasons-responsiveness. I am worried that saying that attitudes are exercises of some form of control does not amount to more than saying that they are responsive to reasons. It is just a very specific use of the word ‘control’ that deviates from our ordinary concept of control. To defend a notion along Hieronymi’s lines requires one to spell out what the talk of ‘direct non-voluntary control over mental states’ adds to the remark that mental states are responsive to reasons. This is a challenge for mental compatibilists more generally.

non-voluntary – can provide this explanation. If one does not provide such a notion of control and an explanation of how it can explain the relevant responsibility, then solving the problem instead requires an explanation of how there can be responsibility without control. It is the latter task this book is devoted to.

Secondly, the problem is formulated in terms of how we can be responsible to the *norms of rationality*, rather than just in terms of how we can be responsible for mental states.³⁴ This makes the problem more pressing, for it locates what is *puzzling* about mental responsibility. It is not *really* puzzling that we are *responsible* for our mental states: for our mental states can be foreseeable consequences of our actions as any other state in the world can be a foreseeable consequence of our actions. What really *is* puzzling is a specific *kind* of responsibility: the responsibility we have in virtue of the fact that our attitudes can be rational or irrational, i.e. that we can respond correctly to our reasons or fail to do so. It is hard to see what form of control could explain *this* form of responsibility. This is the core of the problem. It requires us to understand this form of responsibility, and especially how (or if) it is different from the form of responsibility we have for our attitudes in virtue of indirect control.

2.7 SUMMARY

By thinking about we intuitively lack control over our attitudes (2.1, 2.3) and how we are yet intuitively directly responsible for our attitudes (2.2), we were led to a precise formulation of the problem of mental responsibility. We saw that neither indirect voluntary control nor direct voluntary control are intuitively suited for an exhaustive explanation of our mental responsibility (2.2, 2.4). The reason for this, it turned out, is that we are intuitively responsible to the standards of rationality (2.5). It is this kind of responsibility that is intuitively not to be made intelligible by any form of control we are familiar with.

Thus, what is at issue when we ask how we can be responsible for our attitudes is not just any kind of responsibility, but rather the specific kind of direct responsibility we have to the standards of rationality. I have spelled out the various possible answers one might give in reply to this problem (2.6), and I have pointed out that I will endorse mental involuntarism, i.e., the claim that our responsibility to the standards of rationality does not presuppose control. Yet a big part of my solution will be to contrast our responsibility to the standards of rationality

³⁴ Some authors have already emphasized normativity, rather than responsibility, when formulating the problem. Alston (1988) sees the very possibility of normative epistemology threatened by the absence of doxastic control, Owens (2000) sees the very idea of epistemic rationality threatened by the problem, and McHugh (2014a) describes a conflict between prescriptive doxastic normativity and the absence of doxastic control.

with the kind of responsibility we have to practical requirements. It is only this contrast that will make involuntarism plausible, and it is this contrast that will allow us to see that *The easy solution* is plausible *under a certain reading* – though not the reading the indirect mental voluntarist hopes for.

Chapter 3

What Attitudes Should We Have?

In this chapter, I will first explain what kind of answer I seek when I ask what attitudes we should have. I do this by first reminding us of the distinction between the positions of Clifford and James, on the one hand, and evidentialism and pragmatism, on the other (3.1, 3.2). The question we need to focus on if we want to understand our responsibility to the standards of rationality is the question of what attitudes it is rational to have, which I understand as a structural question asking for the kinds of reasons that are relevant for determining the rationality of an attitude (3.3). I then spell out how thinking about the rationality of attitudes will help us to understand our responsibility to these standards, thereby specifying my approach (3.4). In a final section, I consider the relevance the answer I will give to the question (3.5).

There could be different kinds of answers to the question of what we ought to believe. Consider the following answers:

- (1) You ought to believe what is true.
- (2) You ought to believe whatever is sufficiently supported by your evidence, and refrain from believing whatever is insufficiently supported by it.
- (3) You ought to believe what is (epistemically) rational to believe.
- (4) You ought to believe whatever is best to believe.
- (5) You ought to believe that for which there is sufficient evidence to rationally expect that it is best to believe it.
- (6) You ought to believe whatever pleases your parents.
- (7) You ought to believe what you desire to be true.
- (8) You ought to believe what your parents desire to be true.
- (9) You ought to believe in God.
- (10) You ought to believe anything your teacher tells you.

There are some obvious differences between these statements. First, (1) to (3) are more evidentialist, while (4) and (5), and probably also the rest, are more pragmatist ((9) need not be if we assume that we have sufficient evidence for god; also (11) may not be true on pragmatist's grounds, if we assume that teachers are good sources of knowledge for their students). Secondly, some answers are more *general* than others. In particular, (1) to (5) seem to be more

general than answers (6) to (10). This is reflected in the fact that the first six answers are *prima facie* plausible candidates for answers which may be true independently of person and context. By contrast, we can quite easily imagine situations in which the last five answers are wrong. If we talk about how an ethics of belief should look like, we arguably must look for principles which are as universal as possible – at least if we can expect that the discovered principles can be helpful in applying our ethics to concrete cases so to help us with concrete decisions or with our reasoning about what attitudes to have and how to relate to others, which is (or should be) the aim of ethics.¹ Thus, we would have to decide between answers of kind (1) to (5), not between answers of kind (6) to (10). This does not mean that answers of the latter kind are of no interest. Rather, it seems that we first need answers of the first kind in order to be able to properly assess answers of the latter kind in a systematic way. The ethics of mind this book is concerned with asks for these *fundamental* answers which say what attitudes *all of us* should have *in every possible situation*. That is, I am looking for an answer that is as general as the one Clifford gave (see chapter 1). However, we should first remind ourselves of how the question Clifford and James asked is different from the one we ask in order to understand our responsibility to the standards of rationality.

3.1 CLIFFORDIANISM/JAMESIANISM AND EVIDENTIALISM/PRAGMATISM

As I have argued in chapter 1, Clifford and James were concerned with the question of how to *manage* our beliefs. They were concerned with the norms that guide our belief-managing activities – they argued about which norm fundamentally should guide these activities: norms that require us to always investigate properly so as to ensure (as far as we can) that one's beliefs are properly based on one's evidence, and that thus require us to make ourselves refraining from judgment when our evidence is indecisive, or rather norms that allow us to let our

¹ Another position would be that the aim of ethics is the clarification of ethical concepts, like 'reason,' 'value,' 'goodness,' 'ought,' or 'norms.' This, I think, cannot be ethics' *ultimate* aim. For what is the point of clarifying concepts? The clarification of concepts is itself an ordinary activity we can engage in for reasons that are better or worse. It does not seem to be *intrinsically* valuable to clarify concepts. Rather, our conceptual clarifications should *help us* in some sense. Since ethics, however understood, is guided by questions about what we should do (or how we should be), the conceptual clarifications should help us in thinking about and deciding what to do (and how we should be). One could argue that this is no longer something *philosophy* should be concerned with. Rather, every individual should decide what to do for themselves. However, what is the individual *doing* in deciding what to do and how to be if not weighing their reasons for and against certain courses of action and attitudes, where this weighing assumes certain ethical principles and values? Certainly, ethical considerations are meant for every individual to help them in their decision procedures. Ethics which loses this aim out of sight thereby loses much, if not all, of its value.

‘passional nature’ influence our beliefs sometimes when the evidence is indecisive? We can thus distinguish the positions of Clifford and James with respect to this question:

Question 1: How should we manage our doxastic attitudes?

Cliffordianism: We should always ensure that our beliefs are properly based on our evidence (by, e.g., investigating properly).

Jamesianism: We are sometimes allowed not to base our beliefs properly on our evidence (by, e.g., allowing our passions to influence our doxastic lives).

The phrase ‘to base one’s beliefs properly on our evidence’ includes both to believe what is sufficiently supported by our evidence and not to believe what is insufficiently supported by it.² I have argued against Cliffordianism in chapter 1: it rests on highly questionable empirical assumptions about the harmful effects of epistemic irrationality.³ I thus endorse Jamesianism. It is not mysterious at all how we can be held responsible to the norms Clifford and James are talking about. For these are norms of *managing* our beliefs. These managing-activities are ordinary actions that are subject to the requirements of prudence and morality, which can be placed on us insofar as we can exercise voluntary control in these activities. Discussing Cliffordianism and Jamesianism in more detail will thus not help us in solving any puzzle, for our responsibility to the relevant norms does not give rise to a puzzle.

Evidentialism and pragmatism, by contrast, are usually understood in contemporary debates as answers to a different question, namely

Question 2: What beliefs should we rationally have?

Evidentialism: We rationally should always properly base our beliefs on our evidence.

Pragmatism: We are sometimes rationally allowed not to properly base our beliefs on our evidence.

The difference to the question Clifford and James asked is that the norms of doxastic rationality are *irreducible* standards of belief. To say that someone irreducibly ought to believe something is to say that this ‘ought’ is not reducible to an ‘ought’ by which one expresses an obligation

² I discuss epistemic norms in more detail in ch. 4. We will see that it seems hard to make it intelligible that we always ought to properly base our beliefs on our evidence in the sense I describe here. So if we want to say that it is a norm of epistemic rationality to which we are held responsible that we always properly base our beliefs on our evidence, then we would have to understand the phrase differently.

³ For this critique, cf. also Lindner *fc.* Lindner argues, however, that we can make a good case for Clifford’s position by acknowledging that a belief that is formed in response to practical considerations (i.e., considerations about the belief-state’s value) is not a belief at all. I doubt this premise of her argument, however, for it seems to rule out a lot of paradigm cases of irrational belief.

to perform certain actions, like a requirement to investigate properly. That is, such an irreducible ‘ought’ can be in place even if the person lacks any means to indirectly manage, influence, or control their beliefs through their actions. ‘Basing one’s beliefs on one’s evidence’ is not an activity one performs.

Irreducible ‘oughts’ are to be analyzed in terms of *reasons for belief* rather than *reasons for action*: you irreducibly ought to believe that *p* iff you have decisive reasons to believe that *p*. By contrast, you ought to manage yourself into the state of believing that *p* iff you have decisive reasons to manage yourself into the state of believing that *p*. The dispute between evidentialism and pragmatism is thus about reasons that directly favour attitudes rather than reasons that directly favour actions of managing one’s attitudes. The dispute between Cliffordism and Jamesianism, by contrast, is about the latter type of reasons.

The standards of rationality are the relevant irreducible ‘oughts.’ For to be rational, according to the concept of rationality I adopt here, is to respond correctly to one’s reasons for attitudes. If evidentialism is true, then it is true because all reasons for belief are provided by the evidence, i.e., all of them are epistemic reasons. If there were such things as practical reasons for belief – i.e., considerations that favour a belief by indicating or constituting the practical value or utility of the belief –, then we could conceive of cases where it is allowed to believe for a practical reason.⁴ Pragmatism is thus the view that the rationality of belief is not only a matter of epistemic reasons, but also of practical reasons for belief. While Jamesianism is not committed to the existence of such practical reasons for belief (because it is not a position about reasons *for belief* at all), pragmatism is so committed. This is why I can endorse Jamesianism but deny pragmatism.⁵

3.2 EVIDENTIALISM VS PRAGMATISM: A ROUGH SKETCH

One of the most important arguments in the discussion between evidentialism and pragmatism is Shah’s argument against the existence of practical reasons for belief. We might wonder what

⁴ One might argue that, in principle, we could construe a position according to which all practical considerations for which we can believe are not weighty at all, so that we can never be allowed to deviate from our epistemic reasons even though there are some practical reasons for belief. I will ignore this position due to its extravagance: it is not clear to me why anyone should defend the idea that there are such irrelevant practical reasons for belief.

⁵ As I will discuss in ch. 7, some pragmatists argue that if I endorse Jamesianism, I also need to endorse pragmatism, because there is no clear line between reasons to cause an attitude and reasons for this attitude. I argue that there is such a line, and that pragmatism needs to be understood as being concerned with reasons for this attitude rather than with reasons for managing one’s attitudes. Thus, the distinction between reducible and irreducible ‘oughts’ to have certain attitudes which I introduce here will be reinforced in ch. 7.

other considerations than *epistemic reasons* or *evidence* could be relevant for answering Question 2. According to Nishi Shah, ‘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’ (2006, 481-482). When I consider whether I should believe that it will rain tomorrow, I will check the weather-forecast, rather than considering whether harbouring this belief would, for example, make me depressed. Call this the *transparency principle*: the question about whether to believe that *p* is *transparent* to the question about whether *p*.⁶

Shah argues that only considerations which can function as premises in doxastic deliberation are reasons for a belief. Call this the *deliberative constraint*. However, if doxastic deliberation – that is, deliberation about what to believe – is only possible by thinking about what is true (i.e., about ‘whether *p*’), as the transparency principle claims, then it follows from this principle and the deliberative constraint that only evidence or epistemic reasons are reasons for belief.

If Shah’s argument was sound, then there is no distinction between the question of what we epistemically ought to believe and the question of what we ought to believe (period). For then there would be no considerations relevant for what we ought to believe beyond epistemic considerations, that is, beyond considerations which directly bear on the truth of a proposition.

I will later argue that Shah’s reasoning is sound (part III). There are indeed no non-evidential reasons for belief. The fact that it would be pleasurable to have a certain belief is not a reason for this belief. This, however, does not preclude considerations about the *value* of a certain belief to be relevant for answering ‘what we ought to believe’ in the sense of ‘what beliefs we should ensure we have.’ For answering this question, only reasons for actions are relevant – the transparency principle does not hold for this interpretation of ‘what we ought to believe.’ I explain this meaning of ‘what we ought to believe’ in more detail in chapter 7.1.⁷

Still, other authors think that there are non-epistemic reasons for belief (McCormick 2015; Reisner 2008; 2009; Rinard 2015; 2017). They need to either deny the transparency of ‘whether to believe that *p*’ to ‘whether *p*’ or they need to deny Shah’s constraint on what

⁶ This transparency of doxastic deliberation was first pointed out and labelled that way by Moran (1988).

⁷ My argument against pragmatism in part III will be different from Shah’s in that my reasons for accepting the premises are different from his. Yet my argument throws some light on why the question ‘whether to believe that *p*,’ understood as asking for *irreducible* reasons for belief, is transparent to the question of whether *p*. This is because irreducible reasons for belief are only epistemic reasons, which I will argue for on grounds that are independent from Shah’s argument. Pragmatists rightly observe that the question ‘whether to believe that *p*’ can also naturally be understood as asking for non-epistemic reasons, namely, reasons for actions of belief-management. However, by observing this, they do not establish pragmatism as defined above.

considerations can count as a reason. I discuss such strategies in chapter 7. It might seem that pragmatists are committed to the idea that we can in some sense weigh practical and epistemic considerations when thinking about what to believe. Imagine you have a lot of evidence for *p*. What should you believe if you would, say, get a lot of money for not believing that *p*? According to Andrew Reisner, if there is a way for you not to believe that *p*, there are some thumb-rules by which we can determine whether the epistemic reasons outweigh the practical reasons. For example, he argues that ‘when the pragmatic reasons become strong enough, the evidential reasons cease to play any role at all in determining what one ought to believe’ (2008, 21). In other cases, however, where the advantages we gain by believing against the evidence are not that big, only the evidence counts (*ibid.*, 23). If both evidential and practical considerations are relevant for what we ought to believe, then it makes sense to ask what we ought to believe in light of both kinds of reasons. We might say that it would make sense to ask what we ought to believe, *all things considered*.

However, Reisner also describes a view according to which the evidence does not play any role *at all*, and according to which only the practical reasons we have for a belief are relevant for what we ought to believe (2008, 20). Here, determining what we ought to believe would be similar to determining what we ought to do. We would have to consider different values, like the pleasure a belief might give us and others, or how having a belief will influence our relationships, or the amount of justice the belief could bring into the world. If we could answer the question of what we ought to believe by considering the values we realize by having a belief, the question of what we ought to believe, *all things considered*, would obviously also have a clear meaning: the things to consider would be the values we (probably) realize by believing.⁸ Still, we would not be committed to weigh evidential against practical considerations, because the evidential considerations are relevant to what we ought to believe only insofar as they indicate that the belief that is supported by them would be valuable to have. Such a view is currently defended by Susanna Rinard.⁹

⁸ If you prefer a non-consequentialist framework, you would have to formulate this question in another way which is analogous to what you deem to be the right interpretation of the question of what we should do, *all things considered*. Maybe we do not have to consider the value of the consequences of the action and belief, but rather the intentions or motives or maxims with which we act or believe. As deontologist, we would have to ask: how does deliberation about what to do look like according to this framework? And then we would have to construct deliberation about what to believe as completely analogous.

⁹ Cf. Rinard 2015; 2017. Meiland 1980 is close to such a radical view as well.

3.3 'WHAT ATTITUDES SHOULD WE HAVE?' AS A STRUCTURAL QUESTION

I will understand the question of what to believe (in the sense of Question 2) as aiming at *deciding* between the possible answers I just described. I do not seek an answer to concrete cases under the paradigm of one of these accounts. I do not want to know in the context of this investigations, e.g., whether certain pragmatists would say that we ought sometimes not to believe in human-induced climate change (say, because in these contexts the believe would make us too depressed). Of course, whether we endorse evidentialism, or a moderate form of pragmatism, or a radical form of pragmatism, will also help determine such concrete questions. Yet when I ask the question of what to believe I want to know what role evidence and practical considerations play in determining what we ought to believe. It is a very general, structural question about the norms of belief that demands a very general, structural answer.

Analogously, I want to know how the *kinds of reasons* we can have for attitudes determine what attitudes we should have. If a very sad event happened in your life, but if it would, at the same time, be very bad for you to feel grief right now (say, you have to focus on an important task) – is it then still true that you (rationally) *should* feel grief? Or, if you would survive only if you do not fear a danger you are confronted with (imagine you would be more focussed without fear), is it then still true that you (rationally) *should* fear the danger? After all, it is still danger! More generally, if not having an attitude that seems to be decisively supported by object-given reasons would provide big advantages for you or for others, is it still true that you (rationally) *should* have the attitude?

My account of reasons for attitudes that I defend in part III will imply that you still (rationally) *should* have the attitude. This is because the advantage of having the attitude does not provide you with a reason for this attitude, but only with reasons to *manage* yourself into having this attitude. The only reasons there are for attitudes are object-given. Some questions arise naturally: what is the meaning of the 'ought' of object-given reasons, and what is its significance? The significance, or as I will prefer to say later, its *normative force*, is called into question by the fact that it does not seem to *matter* whether we comply with this 'ought' in cases where we lack any reason to manage ourselves into the attitude we (purportedly) ought to have: nobody seems justified in *blaming* us for, say, not fearing a dangerous object if we thereby could manage to survive. Furthermore, what should we say about cases in which we *ought to manage* ourselves into having an attitude that is insufficiently supported by object-given reasons? In what sense is it still true that we *ought* to have such an attitude if we, at the same time, ought to manage ourselves *not* to have it? These two puzzles – one about blame and

one about conflicting ‘oughts’ – will be spelled out in detail, and partly answered, in part II of this book. Part IV elaborates on my partial answer.

The question of what attitudes we should have only arises because there are situations in which practical considerations and object-given reasons seem to conflict. Sometimes, an attitude that is unsupported by (purported) practical reasons is supported by object-given reasons, and vice versa. Sometimes, there are even strong (purported) practical reasons *against* an attitude which is well-supported by object-given reasons, and vice versa. What attitude should we have in those situations? Here is how we can put the question relevant to my investigation:

Question 3: What attitudes should we rationally have?

Mental Constitutivism: We always (rationally) ought to properly base our attitudes on our object-given reasons. I.e., to be rational is to respond correctly to one’s object-given reasons.

Mental Pragmatism: We are sometimes (rationally) allowed not to base our attitudes on our object-given reasons. I.e., to be rational is to respond correctly to one’s practical reasons, *or* to respond correctly to one’s overall object-given and practical reasons.

As mentioned above when discussing evidentialism and (doxastic) pragmatism can distinguish between two forms of pragmatism: a radical and a moderate one. The radical version of mental pragmatism denies that there are any object-given reasons for attitudes (Rinard 205; 2017). The less radical ones argue that what attitudes we ought to have is a matter of both object-given and practical reasons (McCormick 2015; 2017; Reisner 2008; 2009). This distinction is analogous to the distinction by Reisner I mentioned above between a position according to which evidence does not play any role at all in determining what we ought to believe (or according to which evidence only plays at most an *indirect* role by indicating the value of having an evidentially backed belief), and according to which evidence and practical reasons for belief need to be weighed in some way in order to determine what to believe.

To illustrate how *mental* pragmatism might seem more plausible than *doxastic* pragmatism, consider the attitude of *hope*. According to McCormick (2017), the norms of rationality for hope include that we must acknowledge practical reasons for hope. She argues that when we tell someone in the middle of the desert not to give up hope now, for otherwise we will never make it home (132), we say that it would be *irrational* not to hope, *because* there is a practical reason for hope. Intuitively, the existence of practical reasons for hope might seem more plausible than the existence of practical reasons for belief. For hope, by its very nature, seems to be a state which ‘promises to be able to lift us out of the panics and depressions to

which we are naturally prey and to give us firm direction and control' (Pettit 2004, 160). The nature of this attitude seems to imply that practical reasons are relevant to its rationality. So one might think that there are practical reasons with respect to some attitudes, but not with respect to others.¹⁰

My argument for mental constitutivism with respect to *all* mental states will only be finished after I have argued that forming attitudes for practical reasons requires that one can form attitudes at will, and that we lack such an ability for conceptual reasons. This will be my main argument in part III, but I state it here already to give a first impression of it:

- (1) A consideration can only be a reason for an attitude if we can form an attitude *for* this consideration. (*deliberative constraint*)
- (2) We can only form an attitude for practical considerations if we have direct voluntary control over this attitude. (*pragmatism implies voluntarism*)
- (3) We do not have direct voluntary control over our attitudes. (*voluntarism is false*)

Thus: A practical considerations is not a reason for this attitude.

The first two premises are plausible for the following reasons. If a consideration is a reason for this an attitude, then this must mean that it is somehow and at least sometimes possible to form this attitude *for* this reason. This idea is the essence of Shah's 'deliberative constraint on reasons' (Shah 2006, 484-488), which I already mentioned above when describing Shah's argument for evidentialism.¹¹ If we claim that the pleasure of having a belief can be a reason *for this belief*, this implies that we can form the belief for this reason. Otherwise, the pleasure could just provide a reason to bring the belief about, and could thus only be a reason for an action, but not for a belief.

¹⁰ My label 'mental constitutivism' might seem to be misleading when we consider such attitudes as hope. For arguably, McCormick's case for practical reasons to hope is meant to imply that the constitutive aim of hope is partially practical. Yet for most other mental states, it is not plausible to say that the constitutive aim of these states is practical. Belief, for example, can be distinguished from other attitudes by reference to its constitutive aim of truth, but not by reference to any purportedly constitutive aim of having practically beneficial beliefs. I thus stick with the label of 'mental constitutivism.' It is motivated by the idea that, according to this position, the only reasons there are for attitudes are object-given reasons which are reasons provided by an attitude's constitutive aim (truth, desirability, rightness, etc.). And for most attitudes, this aim is intuitively not that of having an attitude that is beneficial. If practical value was the aim of more than one attitude, then we could no longer distinguish these two attitudes by that aim, and thus the aim could no longer be considered as being constitutive of that these attitudes.

¹¹ I return to the deliberative constraint in ch. 7.2., where I bring Shah's argument on the table again. Cf. also Raz' idea that a reason for something must be capable of guiding it: 'If that R is a reason to ϕ then it must be possible that people ϕ for the reason that R and when they do, that explains (or is part of an explanation of) their action [or, for that matter, their forming or having a mental state]' (Raz 2011, 27, my add).

Premise (2) is plausible because forming mental states for *practical* reasons *means* that you are able to *voluntarily* control your attitude: It means that you can have an attitude just *for* the reason that it is good for you to have this attitude (even in the absence of sufficient object-given reasons for the attitude). This kind of control would be direct voluntary control because it would be the kind of control which we also exercise when we act. Thus, (1) and (2) are intuitively very plausible. I return to them in more detail in part III.

I now turn to sketching my general approach of how to engage with the question of what attitudes we should have in order to provide a solution to the problem of mental responsibility. It will turn out that answering this question is only of secondary interest (though such an answer is an important part of my overall approach). What is of primary interest is to understand the meaning of the rational ‘ought’ by understanding the nature of object-given reasons. We will be in a position to understand how there can be such a thing as mental responsibility if we understand how we are responsible to the standards of rationality – as contrasted to how we are responsible to practical requirements that govern our actions, like the requirements of prudence and morality.

3.4 UNDERSTANDING RESPONSIBILITY BY UNDERSTANDING REASONS: THE APPROACH OF THIS BOOK

I have argued in chapter 2 that the problem of mental responsibility is ultimately a problem about understanding our responsibility to the standards of rationality. This led me to present the question about the content of these norms: are they standards that are merely set by an attitudes constitutive aim, like evidential norms in the case of belief, as mental constitutivism claims, or are they standards (or even requirements) that are also partly practical, as mental pragmatism claims? I have already mentioned, in subchapter 3.3 above, that endorsing mental constitutivism would give rise to a puzzle about the significance of the standards of rationality, which will be at the center of part II of this book. Thus, thinking about the problem of mental responsibility leads us to *another* puzzle about the significance of the standards of rationality. This further puzzle can be expressed, roughly, by asking the following question: Why does it matter at all whether we comply with our object-given reasons if not complying with these reasons does not have any significant consequences? For example, why does it matter whether I believe in, say, the newest celebrity gossip upon being confronted with sufficient evidence that the gossip is true? If we assume that such gossip is of no interest to us, then why should it matter in such cases whether I adopt the belief that is rational given my evidence?

Mental pragmatism has an easy reply to this further puzzle: if it really is of no practical value whether you believe the gossip, then you have no practical reasons to believe the gossip. Pragmatists could thus argue that it is not irrational if you still do not believe the gossip. It seems that evidentialists do not have this option: as soon as you are confronted with sufficient evidence for *p*, you ought to believe that *p*.¹² It does not matter whether it is important or of interest whether *p*: epistemic reasons, and object-given reasons in general, are content-independent.¹³ It is this content-independency that gives rise to the worry that it does not matter *on its own* whether we comply to our epistemic and other object-given reasons. Complying with these reasons only seems to matter when there is something at stake – as it is often not when our attitudes are concerned with trivial contents. Pragmatists seem to be able to capture this by saying that we only ought to believe something if there are not only object-given, but also *practical* reasons to form a rational attitude about the matter at issue.

I think that the puzzle of mental responsibility and the puzzle about the significance of the standards of rationality are interconnected. If we understand the significance of the standards of rationality properly, then we might thereby get a clue as to how we are responsible to these standards. This is the guiding idea of the following investigation. I will thus start with presenting the puzzle about the significance of the standards of rationality in more detail throughout part II. Throughout part II, I assume that these standards say that we ought to respond correctly to our *object-given* reasons: I assume mental constitutivism. This is both because the puzzle arises from this assumption and because my account in part IV will be constitutivist. I will sketch my initial reply to the puzzle of the significance of the standards of rationality by the end of part II (chapter 6.3 and the ‘Summary and Conclusion of part II’ after it), where I explain how this reply can also help us understanding our responsibility to the standards of rationality. The lesson we should draw from the existence of the puzzle is not that there is no such responsibility, but rather that we have to understand the responsibility to rational standards in different terms than we understand our responsibility to the requirements of action, including our responsibility to indirect moral and prudential requirements of attitude-management.

The assumption of mental constitutivism will be questioned by bringing mental pragmatism on the field in part III: in contrast to constitutivism, pragmatism seems capable of

¹² I ignore some complications in formulating an adequate epistemic standard here. I discuss some of these complications in ch. 4. One background condition that must be fulfilled for such an epistemic obligation to be in place is that one attends to the question of whether *p*.

¹³ I take the notion of content independency from Kiesewetter (ms).

providing an easy explanation of the significance of the standards of rationality by saying that it is always practically valuable to be rational. However, I argue in part III that pragmatism fails: there are no practical reasons for attitudes because we cannot believe at will. This will leave us with the challenge of making the significance of the standards of rationality, and thus the nature of our responsibility to these standards, intelligible while being committed to mental constitutivism.

At this point, one might still be puzzled why I think that we can understand the mental responsibility by thinking about the significance of the standards of rationality (and thus about reasons for attitudes). I hope that the idea will become clearer as I proceed. Yet I would like to point out that my approach is not wholly esoteric. Rather, it is inspired by the approach of McCormick (2015). McCormick's book is motivated by the puzzle of how we can be responsible for our beliefs. The first part of her book, however, outlines a theory of *reasons* for belief. It is her pragmatic account of reasons for beliefs that requires her, in the second part of her book, to spell out a notion of doxastic control that is robust enough to allow our beliefs to be subject to the same norms as our actions are.¹⁴ My strategy is similar to McCormick's: I also start out by considering reasons for attitudes. Yet it turns out throughout parts II and III of this book, *pace* McCormick, that reasons for attitudes are to be understood in different terms than reasons for action. It is not clear then, why a theory of control would be required for my own account. Indeed, the theory of responsibility I defend in part IV will make it intelligible that there is no need to think about how we control our attitudes in order to understand how we are responsible to the standards of rationality.

Indeed, I think that the focus on non-voluntary forms of control led us away from what is the central task in understanding mental responsibility: namely, understanding the normative force of reasons. In avoiding a lengthy discussion of forms of direct, non-voluntary control, I make a plea for a shift of focus in debates about responsibility away from control and towards reasons. Putting the conceptual relations between responsibility and reasons at center stage is also meant to motivate more exchange between debates about responsibility and debates about reasons, which often just exist next to each other.¹⁵

Yet the issue of control should not wholly fall from the table of discussion. Control will play a role at two points of my argument. First, in order to see whether we can understand

¹⁴ For my discussion of McCormick's account, cf. ch. 9.4.

¹⁵ To be fair, there are philosophers that discuss reasons and responsibility as deeply intertwined. To be mentioned are Pamela Hieronymi, Miriam Schleifer McCormick, David Owens, and Angela Smith. It will be obvious for the reader who knows their work that the approach spelled out here is sympathetic, or at least motivated, by their approaches.

reasons for attitudes as value-based (in a similar way as many practical reasons are), we need to discuss whether we can believe at will – or so I will argue in chapter 7. I deny that we can believe at will in chapters 8 and 9. Another important issue is whether there are good reasons for explicating a notion of non-voluntary control that are independent from the problem of mental responsibility. I take away one of the most obvious motivations in chapter 12. There I argue that my understanding of mental control as merely indirect can still account for the intuition that we are control our attitudes in a way we cannot control the attitudes of other people, or how we can control non-intentional features of ourselves, like our health, or our headaches. I do so by spelling out a specific form of control that we can only exercise over our own mind, but which still does not amount to *direct* control over our own mind. It is a kind of control that can give rise to specific indirect mental requirements that we can only fulfill with respect to our own mind, and thus demarcates a space of our mind for which we have a specific kind of indirect responsibility.

A guiding idea throughout my investigation is that the significance of a norm expresses itself in the reactive attitudes we show towards norm violations, and being a legitimate subject to these reactive attitudes are constitutive of what it is to be responsible. Violating norms has, according to the account of responsibility I defend, often specific normative consequences that are characteristic of these norms: in the case of violating moral norms, people get wronged as a consequence; in the case of violating rational norms, we often end up impairing our relationships to others, even if we do not wrong anyone by being irrational (though we sometimes do wrong others by being irrational, as I argue in chapter 11). Such normative consequences give rise to sanctions, blame, or criticism. Sanctions, blame, or criticism are, as I will say, *markers* of normative consequences. A normative consequence could be that you harm others, or that you wrong others, or that your relationship to another person is impaired. Sanctioning, blaming, and criticizing are ways of marking such events. If we understand how we mark the normative consequences of the violations of the norms of rationality, then we thereby understand the significance of these norms, and thereby the nature of our responsibility to the norms of attitudinal rationality.

Instead of elaborating on this theoretical framework any longer here, I rely on the reader to trust in me that the approach I sketched is intuitively interesting enough to continue engaging with it. Its fruits will only be harvested by the end of the book, when we see the reply to the puzzle of mental responsibility that results from it. Before beginning our actual investigation, let me comment on the relevance of my position.

3.5 THE RELEVANCE OF MENTAL INVOLUNTARISM

There is an important worry with understanding the standards of rationality as not presupposing control in any way – as I do by arguing for *mental involuntarism* (see chapter 2). The worry is that these standards are wholly uninteresting to us if it is not in our power to comply with them. If the standards of rationality that we are responsible to, according to my position, are such that we are responsible to them even if we lack any control over whether we comply with these norms, then how can they guide us in our life?¹⁶

To see the worry better, consider first how Cliffordianism and Jamesianism are informative for life-guidance. It will make a difference to our lives whether we accept that we should always ensure that we believe what is sufficiently supported by our evidence, or whether we accept that we are allowed to let our passions make us irrational in believing sometimes. Arguably, many arguments over religion might depend on which fundamental principle about such indirect doxastic requirements we accept. While religious people could consistently accept a Jamesian pragmatism so as to justify why they allow themselves to believe in God even in the absence of sufficient evidence for God's existence, critics of religion might base their criticism on the view that it is always morally pernicious to believe against the evidence, together with the claim that there is no sufficient evidence for God's existence.¹⁷ Think, furthermore, of the cases mentioned in chapter 1.1, where irrational beliefs might be quite useful to boost your own self-confidence, or to enhance your relationships with your near and dear.

Given such empirical assumptions, we would, depending on whether we are Cliffordians or Jamesians, first, deem certain strategies of mental self-management, like considering certain evidence, or deceiving oneself in certain contexts, appropriate or inappropriate; secondly, we would *evaluate* and *judge* others who have non-evidentially backed beliefs about themselves or others quite differently. Thus, what we believe about the fundamental questions concerning how we should deal with our own doxastic life will change our own lives. The

¹⁶ That I present such a worry might cause astonishment in contemporary academic philosophy. However, I deem it to be important that our philosophy is concerned with questions the answer to which has an impact on our lives. This is a conviction I adopt from many schools of ancient philosophy. I think that even conceptual clarifications can help us to live a better life. In Kiesel/Schmidt 2019 (German article), we illustrate how distinguishing forms of mental agency can help us in this way, thereby defending the idea of philosophy as an art of living.

¹⁷ Think, for example, of Richard Dawkins' 'Spaghetti monster'-argument, which claims that we are as justified to believe in a Christian God as we are to believe in an almighty, benevolent Spaghetti-monster (Dawkins 2006). It is important to note here, however, that also critics of religion often rely on pragmatic arguments against belief, for example, if they point out that the history of Christianity is full of blood and violence, and take this to be a reason not to be a Christian (which includes not to believe that God exists) (Beckermann 2013, 29-30).

question of whether to be Cliffordians or Jamesians here is thus one which is worth pursuing for (almost) each of us.¹⁸

Yet I have distinguished Cliffordianism from evidentialism, and Jamesianism from pragmatism. The puzzle I am concerned with arises from the fact that we are directly responsible to the standards of attitudinal rationality, rather than to prudential and moral norms of managing our own attitudes. According to mental involuntarism, which I defend, we are responsible to these standards without being in control over whether we comply with them. This responsibility is quite different from our responsibility to the requirements of belief-management Clifford and James were concerned with – at least insofar as the relevant responsibility does not imply that we are in control of whether we comply with the standards we are responsible to. So how can these standards be relevant for us at all?

It is important to note that this worry from relevance is not restricted to mental involuntarism. Mental compatibilists are not immune to it as well. If we exercise a non-voluntary control in believing, desiring, feeling, or intending, then why should these norms be any more interesting? We have a clear grasp on why norms that presuppose voluntary control are relevant to us: it is in our power to comply with them by making efforts of will. Yet it is intuitively unclear what the power is that we are granted by mental compatibilism: what is it that we *do* when we form a belief after we are finished with all our investigating, experimenting, reading, and active thinking?¹⁹ I do not claim that mental compatibilists cannot say anything informative here. However, *prima facie*, it is hard to see how norms to which we can comply only by exercising *non-voluntary* control are relevant for us. More has to be said about the nature of this control.

My reply to this objection from relevance – a reply which is also open to mental compatibilists – has two parts. First, we are sometimes in indirect control of our rationality, and the account I propose will provide an interesting explanation of why we have, at least usually, good reason to *ensure* that we are rational. Secondly, even if we lack indirect control over whether we are rational, our rationality will have normative consequences for our interpersonal relationships, and thus for how to interpersonally evaluate one another. In this way, getting clear

¹⁸ One could object here that this is not obvious if we assume pragmatism. For then maybe it is better not to know the truth about this issue or to have a false belief about what to believe. I think that, even if this cannot be excluded from the outset, to ask oneself what is true here is a good starting point because we can expect that knowing the truth about ethical principles will be valuable for our lives most of the time. If, on reflection, we realize that we would be better off if we had a false belief about what to believe, then we should probably take measures to induce a false belief in us. However, it seems to be improbable that this would turn out.

¹⁹ Cf. ch. 2.1.

about the nature of the standards of rationality, and why we are responsible to them, has significance for our lives.

Let me start with the first point. As I will argue in more detail in chapter 12, we can often intentionally control our rationality by reasoning. Reasoning is here understood as an intentional activity that aims at figuring out what attitudes are correct in light of the other attitudes and information one has. We can intentionally sit down and think things through, and thereby ensure that our attitudes are more rational than before. Furthermore, we can cultivate our disposition towards being rational by more indirect strategies as well. Activities like going for a long walk, or doing some sports, or meditating are likely to have a relaxing influence on our mind, thereby giving us the opportunity to develop more rational attitudes.

My account will imply that – at least in the absence of abnormal circumstances – we have a strong reason, and often ought to, engage in such activities to promote our rationality. This is because I will propose that being irrational often impairs our relationships to others. However, also being *epistemically* irrational (even if the epistemic irrationality is not disrespectful) will usually have the effect that we cannot be trusted in certain ways, as I argue in chapter 4.3. Insofar as we rightly value certain relationships – friendships, romantic relationships, being trustworthy colleagues, roommates, or family members, for example – we also have a reason to ensure that we are rational insofar as our irrationality would impair these relationships. Since having such relationships with others is arguably part of a good life, my account will imply that having a generally rational disposition is of instrumental value to achieving a good life.

However, even if we subtract from these practical reasons to ensure that we are rational, my mental involuntarism would play a guiding role in our lives. For if our relationships are often impaired by irrationality, then we also have a reason to reconsider our relationships to the irrational. This does not mean that we should end all our friendships as soon as we notice that our friends are all irrational from time to time. However, if certain patterns of irrationality show up – say, someone always gets very angry at you for minor faults – you might rationally expect the other person to notice their irrationality, to acknowledge the relationship-impairment produced by it, and to desire to relate to you in a different way. As I will argue throughout chapters 10 and 11, if you acknowledge that your relationship to this person is impaired, you thereby blame them, and thus hold them responsible, even if the blame is not expressed in such responses as resentment or indignation. Such passionate blaming responses might be inappropriate because you might well know that the person could not do otherwise than to be irrationally angry at you (or lacks any kind of indirect control over their attitude, if one has worries

with alternative possibilities). Yet your blame that consists in the marking of the impaired relationship in behaviour and thought is still a rational response.

We can of course always conceive of a case where we should cause ourselves not to blame someone even though the blame would be rational (and this is true even if we think of ‘blame’ merely as an attitude rather than as an expression of the attitude). But this does not show that figuring out what the conditions are that make it rational to blame someone, or to hold someone responsible for their attitude, cannot play a guiding role in our lives. If we get clearer about what kind of blame is rational under what circumstances, this can likely have an effect on the rationality of our overall blaming-responses. Since blaming rationally – i.e., blaming in the right way under the right circumstances and refraining from blaming in the right way and under the right circumstances – is itself be part of healthy relationships, knowing when blame is rational and when it is not will be essential for leading a good life as well.²⁰

3.6 SUMMARY

In the present chapter, I have brought the second main question of an ethics of mind into focus. The question of what attitudes we should have is a structural question concerned with what kind of reasons determine the rationality of an attitude. It is to be distinguished from the question of how to manage our attitudes, with which Clifford and James were concerned with (3.1). We can distinguish between mental constitutivism, which claims that only object-given reasons determine the rationality of an attitude, and mental pragmatism, which claims that also a type of state-given reasons – namely, practical reasons – determine the rationality of an attitude (3.2, 3.3). This book will defend mental constitutivism in part III on the grounds that practical reasons presuppose that we can believe at will, and that we cannot believe at will. I have furthermore spelled out the approach of this book (3.4) and its significance (3.5). A significant part of this book is devoted to understanding the status of the ‘ought’ of rationality (under the constitutivist reading of ‘rationality’ outlined in chapter 0.2.4) rather than with the question of what kinds of reasons determine an attitude’s rationality. I thus approach the question of how we are responsible to the standards of rationality by considering the status of the standards of rationality. If we understand the status of these standards, then we can understand the nature of the relevant responsibility. I begin this approach in the next chapter by spelling out the threat from mental nihilism. Finally, I have argued that a mental involuntarist response to the problem of

²⁰ For a helpful recent discussion of the ethics of blame, cf. Coates *fc.*

mental responsibility can have significance for how we lead our lives by being relevant to how we should cultivate our mind as well as for how we rationally relate to one another and thereby develop healthy relationships.

Part II

***Reasons for Attitudes, Blameworthiness,
and Cases of Conflict***

Introduction to Part II

The easy solution to the problem of mental responsibility claims that responsibility for attitudes is completely analogous to responsibility for consequences of our actions. We saw in chapter 2 that the problem of mental responsibility is a problem about our responsibility to the standards of rationality: it is about understanding the status of standards that say (roughly) that we ought to believe what is sufficiently supported by one's evidence, or to desire what one has decisive reasons to regard as desirable, or to fear what one has decisive reasons to regard as dangerous, or to intend those actions one has decisive reasons to perform.¹

The problem was that, in order to understand how we can be responsible to such standards, it seems that we need to assume that there is a form of control we can exercise over our (or in) believing, desiring, feeling, and intending; and yet it is hard to see what kind of control could possibly explain why we are responsible to the standards of rationality: the familiar voluntary control we exercise when we act did not seem to be suited for this explanatory task.

There is a radical way of establishing *The easy solution*: we could endorse *mental nihilism* (cf. 2.6). That is, we could just react to the absence of direct mental control by endorsing the claim *that there are no reasons for attitudes*: there are no reasons for or against believing something (though there are still reasons for or against *managing* our beliefs through our actions), and there are also no reasons for or against desiring, feeling, and intending (though there are still reasons for *managing* our mind). If there are no reasons for attitudes, then there surely is no responsibility for our (non-)compliance with such reasons left to explain.

As an initial motivation for this position, one could point out how absurd it is to tell someone to just *be* in a specific state of mind, like believing, desiring, etc. People cannot just *be* in states, they can only *do* things in order to bring themselves into states or do things in order to maintain states. So, what, one might ask in philosophical astonishment, is the point of standards that directly govern attitudes rather than actions?

I think these standards have a point. But I think that their task is quite different from the task of *practical requirements* that govern our actions. I can be required not to steal your purse. And I can be required to help you if I see that you are in severe need of help. If I steal your purse or do not help you, you might, in response, resent me, or feel indignant about me,

¹ Remember that I adopt a reasons-based conception of rationality (cf. ch. 0.2.4). The challenge I will spell out in this part of the book is calling into doubt the status of object-given reasons as *reasons*. Defending the idea that we ought to be rational (in a sense) will, in my terminology, mean that one defends the idea that one should (in a sense) respond correctly to one's object-given reasons.

and demand me to return your purse, or demand me to help you now, and you might expect an apology for what I did. When I realized that I have wronged you, it would make sense for me to feel guilty. These kinds of responses to my wronging you would be *rational* responses: in light of what I did, it *makes sense* for you (or me) to so respond to me.

By contrast, if I come to believe that candidate X will win the next elections because the flight of the birds gave me a sign, it does not seem appropriate to *resent* me, and I do not need to apologize to anyone because I did not *wrong* anyone with my belief. I was just irrational. At other times, we might wrong someone with our attitude: I might be racist towards you. If you find out and make me aware of my racism, and I realize my attitude as problematic, then I owe an apology to you, whether or not my racism manifested itself in voluntary conduct. Yet it seems, at least intuitively, that I do not owe you this apology *in virtue of my irrationality*. The problem here does not seem to be that I did not believe what my evidence supported (even though that might be the case), but just that I had a disrespectful attitude towards you.²

This initial contrast between requirements of action and standards of (attitudinal) rationality will, for now, suffice to point out the direction of where my overall argument in this book is heading towards. I will argue, especially in chapter 6 and the subsequent ‘Summary and Conclusion of Part II,’ that mental nihilism is right in pointing out that there is no such thing as requirements to believe, desire, feel, or intend that have analogous normative consequences as the requirements of prudence and morality.³ This fact is what explains our intuition that it is pointless to tell someone that they should just *be* in a mental state. But even if attitudes cannot be directly required in the way actions can be directly required, it is still true that there are standards of attitudinal rationality to which we are responsible. The significance of these standards – or as I will prefer to say, their normative force – is to be understood in different terms than the normative force of requirements of voluntary conduct.

Part II provides the initial motivation for the specific way I distinguish the standards of rationality from requirements of action. I will do this by *exploring* a case for mental nihilism. I will spell out two main arguments in favour of nihilism and explain, as well as partially defend, my own evaluation of these arguments. Properly understood, these arguments do not establish nihilism, but rather a largely ignored ambiguity in our expressions of ‘ought’ and ‘reasons’ that calls for a better understanding.

² Note that this is just how it seems *prima facie*. The criticism might ultimately well be a specific form of rational criticism, as I suggest in chapter 11. Cases of racist attitudes and other irrational attitudes that commonly give rise to emotions like resentment will be central for my discussion in chs. 10 and 11 of part V.

³ On the notion of ‘normative consequences,’ cf. ch. 3.4.

According to the account I spell out, rationality sets a standard according to which we are evaluated as more or less rational or irrational, and this evaluation has specific normative consequences. The idea that we are *required* to have certain attitudes gives rise to the confusion that we need to understand how we can be in direct control of our attitudes in order to understand why we are responsible for them. By contrast, the idea that rational norms are mere standards of personal evaluation with specific normative consequences for our interpersonal relations does not give rise to this idea of direct control. We can rest content with the thought that we are responsible to these standards whether it is in our power to comply with them or not.

Let me finally put part II in the context of the parts to come. Part III will consider an objection to my distinction between standards of rationality and requirements of action. The objection is coming from mental pragmatism, i.e., the idea that the standards of rationality do not merely provide us with object-given reasons, but also with a type of state-given reasons: practical reasons for attitudes, i.e., reasons that show the attitude to be valuable to have. According to pragmatism, we can form attitudes for the same reasons for which we can act, and can thus be said to be subject to practical requirements to believe, desire, feel, and intend as we are under practical requirements to do and not to do certain actions. I deny pragmatism by arguing that pragmatism implies the ability to believe at will, and that there is no such ability. Yet I will spell out a non-traditional and uncontroversial form of pragmatism that merely presupposes indirect control, and that is not in conflict with mental constitutivism (cf. chapter 3 on mental pragmatism and constitutivism).

Part IV will then provide substance to the distinction between the standards of rationality and requirements of action by explicating the distinction via two different kinds of responsibility and blame. An account of the normative force of the standards of rationality will be given in the *outlines*, for spelling out the details of such an account is not necessary for providing a solution to our problem (hence the title of the book). Yet a full account will be necessary for an all-encompassing understanding of the normative force of rational standards – a task I will spell out in this book, but not pursue to the end. I leave it for future investigations.

I will, as I did in the previous chapters, start off again with a focus on the attitude of belief, before I discuss whether we can generalize the discussed argument to all mental states. I do this because this will make things simpler from the outset. Bringing in other attitudes than belief will complicate things, and I will discuss some of these complications in chapter 5.

Chapter 4

The Argument from Doxastic Blameworthiness

I will start out with an initial puzzle about the normative force of evidence: it seems as if failing to correctly base one's beliefs properly on one's evidence can never, by itself, make a person blameworthy (4.1). From this I construct a *prima facie* plausible argument to the conclusion that there are no reasons for belief: the argument from *doxastic* blameworthiness. I then focus on the first premise of the argument which will provide me with some opportunity for defending the connection between reasons and blameworthiness that is both central to the argument and to my own account (4.2). This connection will throw some first doubts on the claim of anti-responsibilism, i.e., the claim that we are subject to the standards of rationality but not responsible to them. I will then explain my strategy for objecting to the second premise – a strategy that will accompany us throughout the rest of the book (4.3). I end the chapter by reconsidering *The easy solution* as it is endorsed by consequential conceptions of doxastic responsibility and reject these conceptions, thereby providing an argument against indirect mental voluntarism, i.e., the claim that we can explain why we are responsible to the standards of rationality by reference to indirect control (4.4). The last subchapter summarizes the results (4.5).

4.1 EVIDENCE AND BLAME

I present an initial puzzle about the normative force of epistemic reasons: we do not seem to be blameworthy for non-compliance with epistemic norms (4.2.1). This gives rise to a dilemma for the idea of epistemic blame (4.2.2), as well as to a powerful argument against the normative force of evidence, which thereby presents us with the threat of doxastic nihilism (4.2.3).

4.1.1 THE CLUTTER OBJECTION AGAINST EPISTEMIC NORMS

Gilbert Harman (1986, 12) famously points out that endorsing the following standard would lead to 'cluttering one's mind':

(ER) One ought to believe everything that is sufficiently supported by one's evidence.

Upon reflection, (ER) seems not plausible. My current belief-stock implies the proposition that [I am sitting in my office *or* the moon is made of cheese *or* Hitler is alive *or* there is no human-

induced climate change *or ...*]. There is no doubt that this disjunctive proposition is true right now while I am sitting in my office, no matter how many odd (but meaningful) claims we add. It is true because the first claim, that I am sitting in my office, is true. At the same time, it seems that I do not always believe this disjunctive proposition while I am sitting in my office. Most importantly, I would neither be irrational nor blameworthy (or criticizable) in any sense for *not* believing such disjunctive propositions. It follows that there is no unconditional norm to believe everything that is sufficiently supported by my evidence. This far, contemporary epistemologists agree.¹

Yet philosophers often want to modify (ER) to a more plausible claim by spelling out under which conditions our evidence requires us to believe a proposition. They thus include *background conditions* for when our *sufficient evidence* provides us with *decisive reasons* to believe, i.e., conditions under which sufficient evidence makes it the case that we *ought* to believe a proposition. In this vein, Benjamin Kiesewetter (2017, 184-5) responds to Harman's clutter-objection by proposing that the central standard of theoretical rationality is to believe what one's evidence sufficiently supports *if one attends to this evidence*. According to this proposal, if I attend to the fact that I have sufficient evidence for a specific disjunctive proposition, then I would be irrational, and thus criticizable, if I do not come to believe it. Thus, I *ought* to believe it.

Yet we can reasonably wonder whether Kiesewetter's modification will do. Steglich-Petersen (2011) argues that epistemic reasons do not establish *by themselves* what we ought to believe. And this, I take it, is independent of whether one attends to them. I may have strong evidence in favour of *p* to which I attend without it being true that I ought to believe that *p*:

Suppose, for example, that the subject matter is whether there is an even number of dust specks on S's desk. Let us also suppose that S has excellent evidence, and thus epistemic reason to believe in the sense defined, that there indeed is an even number of dust specks on his desk. In spite of this epistemic reason, it does not seem to be the case that S ought to form the belief that there is an even number of dust specks on his desk. It may be that S as a matter of fact cannot avoid forming that belief, since we are psychologically disposed to form beliefs that are supported by consciously considered evidence. But it is

¹ One might doubt that I do not believe the disjunctive proposition. For if I was asked whether I believe it, and I understand the content of the proposition, I will reply that I do believe it. However, even if one thinks that we believe all those disjunctive propositions, one will agree that we do not believe all the implications of our current belief-stock, like certain mathematical or logical implications that are just too hard to figure out. That we do not believe those implications does not make us blameworthy or criticizable in any sense. Furthermore, we *can* imagine a case where I fail to believe such weird disjunctive propositions. Why on earth, we might ask, should any reasonable person care about this so much as to regard me as blameworthy for not believing them?

nonetheless not the case that S ought to form that belief. If S failed to form the belief, we wouldn't fault him or regard him as normatively worse off for that reason (Steglich-Petersen 2011, 23).

One might object that two levels of evaluation get confused here: practical and epistemic. Even if S is not necessarily blameworthy from a practical perspective when he does not believe on the basis of sufficient evidence, he is, according to this objection, still blameworthy *from an epistemic perspective*. However, it is important to see here that certain emotional reactions that are usually conceived of as forms of blame – like resentment or indignation – seem to be inappropriate towards the person in Steglich-Petersen's example. Why should we *resent* S for believing some utterly unimportant truth nobody cares about? Surely, S should not feel *guilty* about failing to believe it either. Furthermore, it is unclear what, if any, consequences S's purported epistemic failure should have for future relationships with other people who know about this failure.² Being epistemically irrational only *once* does not seem to have any effect on, for example, whether one is a trustworthy source of information as long as this single case of irrationality is clearly an exception.³ So, in what sense then, is S *blameworthy*?⁴

I will call any form of blame that arises *merely* from the fact that someone fails to properly base one's doxastic attitudes (belief, disbelief, or suspension of judgment) on one's evidence *epistemic blame*. We can reasonably ask whether there is such a thing as epistemic blame. Until now, it seems that *merely* failing to believe what one's evidence sufficiently supports does not give rise to epistemic blame. Rather, it seems that it at least has to *matter* in some way that one believes what one's evidence sufficiently supports in order for such a standard to be in place. Whether it matters, however, seems not itself to be a matter of one's evidence, but rather a matter of *practical* considerations that show it to be important that one believes what one's evidence sufficiently supports. If we end up blameworthy only in cases with such *practical* stakes – for example, when we fail to believe in human-induced climate change or have racist beliefs – then this blame seems to be no longer *epistemic* blame.⁵

² I here have in mind Scanlon's (2008) concept of blame, to which I return in chapter 11 in more detail.

³ It would thus not be appropriate to reduce epistemic trust in the person. Reducing epistemic trust is the essence of epistemic blame, according to Kauppinen (2018).

⁴ One could argue that S is blameworthy from a practical perspective if one thinks that believing what is not sufficiently supported by one's evidence is always *morally* blameworthy (Clifford 1877). However, this is not only highly questionable (Schmidt 2017), it also does not change that non-compliance to epistemic reasons does not make you blameworthy. Rather, it would be your non-compliance with moral reasons that makes you blameworthy.

⁵ In a recent paper, Kiesewetter (ms) acknowledges that there is another, probably more difficult, problem for the normativity of rationality than Harman's clutter objection. This further problem arises from the question whether we have a *reason to care* about being rational. While Harman's clutter-objection can be met, according to Kiesewetter, with the attending-condition for requirements of theoretical rationality, this condition cannot explain why we have a reason to care about being rational. Steglich-Petersen's dust-speck case might be

4.1.2 A DILEMMA FOR EPISTEMIC BLAME

Note that Harman's clutter-objection only gets the grip on us if we assume some conceptual connection between reasons and blameworthiness (or criticizability). We assume that, at least normally, we are blameworthy for failing to respond correctly to our epistemic reasons. Since we are not blameworthy in Harman's (or, for that matter, Steglich-Petersen's) case, we conclude that we did not fail to respond correctly to our epistemic reasons in these cases. Rather, we just did not believe what was sufficiently supported by our evidence. Thus, there is no epistemic standard that we ought to believe everything that is sufficiently supported by our evidence: believing what sufficient evidence supports is not always the response that is decisively supported by our overall reasons for belief.

Instead of opting for pragmatism, we might respond to Harman's clutter objection by modifying the epistemic standard so that its violation more plausibly gives rise to epistemic blame. First, one might propose background conditions for *when* we are required to believe what our evidence sufficiently supports. These background conditions must fulfill two criteria:

- (a) they must make it plausible that the subject is, at least normally, blameworthy when they are fulfilled;
- (b) they should not render the norm a *practical*, rather than an *epistemic* norm.

Call (a) the criterion of *significance*, and (b) the criterion of *content*.

Secondly, one might argue that, although we are never blameworthy merely for failing to believe what our evidence sufficiently supports, there are other cases in which we fail to respond correctly to our epistemic reasons. Specifically, we might defend the following standard of epistemic rationality:

(ER*) One ought not [to believe what is not sufficiently supported by one's evidence.]

In what follows, I will reject both options before I construct the argument for nihilism which is based on the impossibility of epistemic blame, thus posing a challenge for the very idea of reasons for belief and epistemic rationality.

understood as exactly raising this question about reasons to care about being rational. Kieseewetter's solution to this new problem about reasons to care builds on Wedgwood's (2017) idea that rationality is a virtue. Since we always have a reason to care about being virtuous, we always have a reason to care about being rational. This is not in conflict with the account of the normative force of reasons for attitudes which I defend (cf. chs. 4.3, 6.4, part IV). Rather, my account can be read as an *elaboration* on a proposal along Kieseewetter's lines.

Consider the first option. It seems that Kieseewetter's reply to the clutter-objection fails because the proposed background condition does not fulfill criterion (a). There are cases where we attend to sufficient evidence but where it would not make much sense to blame us if we, for whatever reason, do not believe what the evidence supports (as argued in 4.1.1). It just does not *matter* whether we do or do not believe it. Yet if we propose a background condition that implies that it *matters* whether we comply with the evidence, we seem to end up violating criterion (a): if the norm is only in place when it *matters* whether we properly base our beliefs on our evidence, then the norm seems to be no longer a purely *epistemic* norm. It thus seems that there is no background condition on an epistemic norm to believe anything that is sufficiently supported by one's evidence that fulfills both (a) and (b). The defender of epistemic blame is in a dilemma.

Let us provide this dilemma with additional support by considering a background condition that does not fulfill criterion (b). Steglich-Petersen (2011) does not endorse a background condition along the lines of the attending-condition. Rather, he concludes from his thought experiment with the dust-specks with the following analysis of reasons for belief:

Necessarily, if S has all-things-considered reason to form a belief about *p*, then [if S has epistemic reason to believe that *p*, S ought to believe that *p*] (2011, 24).

I take it that 'epistemic reason' must be read as 'sufficient evidence for *p*.'⁶ The conditional then states that one should only believe what one's evidence sufficiently supports *if one has an all-things-considered reason to form a belief about p*. This is the proposed background condition for the epistemic standard.

Steglich-Petersen's 'all-things-considered reason to form a belief about *p*' to be understood as a reason for an action prior to the actual 'formation' of the belief (see esp. Skipper/Steglich-Petersen *fc*).⁷ This 'formation' might be the action of *thinking about whether p*: I may have more or less reason to think about whether something is true. I have some reason to think about whether there will be nice weather during the next days, but I have no reason at all to think about the newest celebrity gossip (rather, I have reason to avoid such thinking). Thus,

⁶ Steglich-Petersen says that 'by "epistemic reasons", I shall refer to considerations that count in favour of holding a belief in a particular proposition solely on account of that (potential) belief's epistemic properties. [...] [B]y "epistemic properties" I shall generally mean properties that are in some way relevant to whether or not the proposition believed is true. On this very general definition, being evidentially supported is the most obvious candidate for an epistemic property, but I shall not exclude other candidates in advance' (2011, 17). What I mean by 'sufficient evidence' can be understood as an instance of the property of being evidentially supported.

⁷ For two senses of belief-formation, see ch. 8.2 of this book as well as Schmidt 2016. It is clear that the reason for forming a belief about *p* cannot be itself an epistemic reason, because epistemic reasons favour a specific proposition – they favour that *p* or that not-*p*, but not 'forming a belief about *p*.'

the doxastic Ought (i.e., the sentence ‘S ought to believe that p’) in Steglich-Petersen’s analysis is conditional on a reason for an action. His analysis implies that whenever we ought to believe that p, we have a reason to bring it about that we have a (true)⁸ belief about p. Since we only have a reason to bring this belief about when it *matters* whether we bring this belief about, Steglich-Petersen’s proposal does a good job fulfilling our criterion of *significance*. Yet his proposed background condition, and thus the proposed epistemic standard, is no longer *epistemic* because it includes a reason for an action. His proposal thus fails to fulfill our criterion of *content*.

To find a way out of the dilemma, the proponent of epistemic blame would have to argue either that we can be blameworthy even though it does not *matter* whether we comply with the norm of belief she proposes, or that there is a background condition which implies that it matters whether we comply with the norm of belief but which is not a *practical* background condition. Both options seem to be hard paths to go. We might instead want to reject the idea that we are ever *epistemically* blameworthy. It now seems a promising option to reject the idea of epistemic blame.

We might look for another purely epistemic standard than (ER). We might instead claim that (ER*) is the fundamental epistemic standard. This leads us to the second strategy for the defender of epistemic blame. If we *lack* sufficient evidence for a proposition and yet we *still* believe in it, doesn’t that make us blameworthy? It does not seem so. While the problem with a norm *to believe everything* that is supported by our evidence faces the problem that it would require us to needlessly clutter our minds, the norm *not to believe anything* that is *insufficiently* supported by our evidence faces the reverse problem: it would require us *not* to have a lot of beliefs we, it seems, have no reason to give up. For example, why should I give up evidentially unsupported but beneficial beliefs? We often overestimate our own abilities or the virtues of our significant others. Arguably, this promotes our self-confidence or our relationships. It seems that these beliefs are blameless as well. Thus, a general norm not to believe what is insufficiently supported by one’s evidence seems too *exclusive*, while a norm to believe anything that is sufficiently supported by our evidence seems too *inclusive*.

Furthermore, cases of trivial belief pose a similar problem for (ER*) as they do for (ER). Take, again, any trivial belief where it does not matter whether I have it. What if I believe, in absence of any evidence, that there is an even number of dust-specks on my table? Assume

⁸ In his later works, Steglich-Petersen accepts that the reason to form a belief about p must in fact be a reason to form a *true* belief about p (see esp. Skpper/Steglich-Peteren fc).

that this belief does not have any *further* consequences on my overall rationality.⁹ Why should it make sense for anyone to blame or criticize me for this belief? Again, it seems that, if it does not *matter* in an instance whether we comply with a norm saying that we should not have beliefs that are not sufficiently supported by our evidence, then we are not blameworthy for non-compliance. However, as soon as it *matters*, the blame we are subject to is no *epistemic* blame. If we want to defend the idea of epistemic blame, then we face the same dilemma no matter what version of the norm to properly base one's doxastic attitudes on one's evidence we choose.

I think the dilemma for epistemic blame constitutes a serious challenge. I will ultimately accept that there are reasons for belief, and thus I will deny nihilism. Since, as will become clear in what follows, I think that this requires me to make sense of a notion of epistemic blame, I am committed to the idea that the challenge for the defender of epistemic blame (and thus for me) can be met. Meeting this challenge should, I argue, make us reconsider the concept of blameworthiness (I make some initial remarks about this in chapter 4.3 below). Yet for the present purposes, we should accept that there is a good case to be made for the impossibility of epistemic blame, which arises from an intuitive grasp of the notion of what it is for a norm or reason to matter in such a way that non-compliance can make a person blameworthy. Let us thus accept this challenge for the possibility of epistemic blame, spell out a bit more clearly how it poses a challenge to the very idea of reasons for belief, and then see where it would lead us, and how we could react to it.

4.1.3 AN ARGUMENT FOR DOXASTIC NIHILISM

It is easy to see how the denial of epistemic blame can give rise to an argument for nihilism about reasons or norms of belief. Take the phrase 'to properly base your belief on your evidence' to refer to the condition on epistemic norms expressed in (ER) or (ER*) above, and, if you want to, include an additional attending-condition along the proposal of Kieseewetter:

The argument from doxastic blameworthiness

- (1) There are reasons for belief only if we can be blameworthy (or criticizable)¹⁰ merely for failing to properly base our beliefs on our evidence.

⁹ See ch. 1.1 on why I think that there are cases of isolated believing against the evidence where there are no consequences for my overall rationality. See also Schmidt 2017, 609.

¹⁰ I will use the notion 'blameworthiness' in what follows for the sake of simplicity.

- (2) We cannot be blameworthy (or criticizable) merely for failing to properly base our beliefs on our evidence.
- (3) Thus, there are no reasons for belief.

According to the view expressed in (3), evidence would gain its *apparent* status as a reason only from the fact that following the evidence normally furthers our (legitimate) *practical* interests.¹¹ So, even if evidence does not provide us with reasons for belief, it *appears* as if it provides us with such reason, because we often have reason to follow the evidence in order to gain *valuable* true beliefs – most importantly, beliefs about truths that *matter* or which we have *reason to care about* (rather than all true beliefs). If we are nihilists, it seems that we are committed to this practical-instrumental approach to epistemic normativity.¹²

Premise (1) is *prima facie* plausible only if we assume that the only reasons for belief there could be are provided by our evidence. If there were practical reasons for belief – i.e., reasons for belief that are provided by the value of having the belief –, then we might be blameworthy whenever we fail to conform to these practical reasons for belief. Premise (1) would be false: there would be reasons for belief (namely, practical reasons) even if we would never be blameworthy merely for failing to properly base our beliefs on our evidence – rather, blameworthiness could arise due to the violation of practical doxastic norms. For now, let us assume that there are no practical reasons for belief and evaluate the plausibility of (1) in light of this assumption.¹³

I take (1) to be plausible because it states a very *minimal* conceptual connection between reasons for belief and blameworthiness. It just claims that it must be *possible* to be blameworthy for non-compliance with epistemic reasons. If it is not possible, evidence does not deserve its status as a reason. As pointed out in 4.1.2, Harman's clutter-objection would not get any grip on us if we would not assume that failing to comply to an epistemic norm is supposed to make us, at least normally, blameworthy (in some sense). If we would not assume this, then we could just accept that by not drawing all the implications from our current beliefs we fail to believe what we ought to believe: we would just constantly violate an epistemic norm, but that would not have any further significance. However, the normative force of this

¹¹ This is different from a position put forward by Owens (2017b), who claims that epistemic norms gain their authority from fact that it has practical value to be *subject* to such norms.

¹² This approach is *one* version of what Skipper and Steglich-Petersen (fc) call *epistemic instrumentalism*. I return to an objection to their account of doxastic normativity on the grounds that it implies doxastic nihilism in ch. 4.2.3 below.

¹³ As I have noted before, I will argue against practical reasons for belief in part III. See also ch. 3.3 for my general line of argument for this evidentialist view.

‘ought’ would then be mysterious: why comply with this norm if no one can hold us legitimately accountable for non-compliance? The norm would at best have the force of the norms of etiquette: we can reasonably ask why we have a reason to comply with the norms of etiquette in a given situation, and there are cases conceivable in which we ought not to comply with them.

Epistemologists propose background conditions on epistemic norms, like Kieseewetter’s attending-condition or Steglich-Petersen’s reason for forming belief about an issue, because they want to make sense of the *significance* of these norms: they want to explain why it *matters* to us whether we comply with the norms, why we can be blamed or criticized if we fail to comply with them. The whole project of spelling out a notion of epistemic blame is, I take it, to understand the normative force of evidence: of why it *matters* to comply with epistemic norms.¹⁴

Yet most importantly, even if we reject the idea that we need to be *always* blameworthy or criticizable for failing to comply with epistemic norms in order for there to be reasons for belief, this would not refute (1). For according to (1), it need only be *possible* to be blameworthy for non-compliance: there need to be cases conceivable in which we are blameworthy *merely* in virtue of the fact that we do not believe what our evidence sufficiently supports (or *merely* in virtue of the fact that we believe something that our evidence does not sufficiently support) in order for there to be reasons for belief.¹⁵

Note that this very weak claim (1) poses a serious problem for the position that I called anti-responsibilism in chapter 2.6. Anti-responsibilists claim that, although there are reasons for attitudes, we are not responsible for complying with them. However, if we were *sometimes* blameworthy for failing to comply with the standards of rationality, then this implies that we are responsible for complying with our reasons for attitudes. It thus seems that anti-responsibilists must, like nihilists, deny that we are ever blameworthy for failing to be rational. Yet according to (1), that would make doxastic anti-responsibilism collapse into doxastic nihilism. Thus, a very weak and well-motivated claim about a connection between blameworthiness and reasons, like (1) already poses a serious problem for anti-responsibilism.¹⁶

¹⁴ A recent account of epistemic blame which is clearly motivated in this way is Kauppinen (2018).

¹⁵ These are just two intuitive examples of epistemic norms. As pointed out above, we might add a attending-condition to the norm. More generally, we might add any condition that does not render the norm practical.

¹⁶ My argument in ch. 11 that answerability is a form of responsibility will further undermine anti-responsibilism. For being answerable means to be liable for inquiries for one’s reasons for an attitude, and if one is unable to provide adequate reasons to back one’s attitude, one might be blameworthy, and thus responsible, for failing to be rational.

One might object at this point that, if (1) is such a weak claim, then (2) needs to be more controversial in order to reach the controversial nihilist conclusion. Yet we can make a strong case for (2) with Steglich-Petersen's dust-speck example. It is not straightforward at all in what sense S (the example's protagonist) is blameworthy. Note, furthermore, that although it is just one case, it has a *general structure*: S does not believe some unimportant proposition p even though S has sufficient evidence for p (to which S attends). That we *can* be blameworthy *merely* for being epistemically irrational – i.e., non-(2) – seems to imply that we sometimes *are* blameworthy in cases of this general kind. For it is only in these cases that we have isolated any non-epistemic factors that could also explain S's blameworthiness in similar cases (like the factor that S violated his practical obligation to *investigate* into a certain topic, or that S does not believe something that would be *morally good* to believe, or that S has a belief that is *disrespectful* to a person). If there is never any blameworthiness left after we have isolated these non-epistemic factors, then it seems that there is no such thing as epistemic blameworthiness. In arguing against (2), we cannot point to cases like disbelief in human-induced climate change or racist beliefs, because they obviously have a moral or practical connotation – it matters for our future actions, their effects on others, as well as for our relation towards others whether we have such attitudes. It thus seems that these cases cannot underwrite the idea of *epistemic* blame. It is hard to see how any other case which both shares the structure of the dust-speck case *and* in which the non-epistemic factors have been isolated could imply that S is blameworthy.

Thus, as long as believing that p is unimportant, or does not matter, or we have no reason to care about it,¹⁷ it seems that we cannot be blameworthy for not believing p. Since evidence only indicates that something is true, and not that having the true belief *matters*, failing to believe what one's evidence supports (or failing not to believe what is not supported by one's evidence) can never, *by itself*, make us blameworthy. It seems as if non-epistemic considerations need to come into play in order to make it plausible that we are blameworthy. It follows that there is no such thing as epistemic blame – and consequently, given (1), no reasons for belief.

It thus seems that we should conclude that there are no epistemic reasons for belief. That may sound unsettling indeed. To avoid this unsettling conclusion, we might try to doubt either (1) or (2) by looking closer at the concept of *blameworthiness*. We also might wonder

¹⁷ Choose the phrasing you favour at this point, because we are still working with an intuitive understanding of these terms here. Thinking about them in detail will be part of my general strategy for defending the normative force of object-given reasons.

about the phrasing ‘blameworthy *merely* for failing ...’ employed throughout the argument from blameworthiness. And we might try to argue that in cases of racist beliefs and disbelief in human-induced climate change, we might find a form of criticizability that is *grounded in the fact* that the subjects violate purely epistemic norms, rather than in non-epistemic considerations.

I think these are promising strategies, and they will ultimately show where the argument from blameworthiness must have gone wrong. I will postpone a closer look at such strategies.¹⁸ For the remainder of this chapter, I want to spell out in more detail why considerations related to blameworthiness pose a serious threat to the very idea of mental normativity.

To do this, I will first (subchapter 4.2) address some initial worries one might have with premise (1) of the argument from doxastic blameworthiness. Since these worries do not bring out any flaws of the argument, I thereby support my claim that premise (1) is uncontroversial.

4.2 SOME INITIAL WORRIES WITH THE ARGUMENT FROM DOXASTIC BLAMEWORTHINESS

I will first consider a potential counterexample against a close connection between reasons and blameworthiness as stated in (1): children and animals who can act for reasons, but who cannot be blameworthy. I reject this as a counterexample against (1) by specifying the meaning of (1). I then consider three positions that might be considered as being in conflict with (1), but I argue that they are not, thereby explicating (1) further: I consider objectivism about the meaning of ‘ought’ and ‘reasons,’ epistemic instrumentalism, and epistemic permissivism. I conclude that (1) is in fact as plausible as I presented it and that we thus need to take a closer look at (2).

4.2.1 A WORRY FROM CONSIDERING CHILDREN AND ANIMALS

One might object to premise (1) by pointing out that children or some non-human animals can act for reasons but *cannot* be blameworthy if they act against decisive ones, because they are not responsible.¹⁹ Analogously, some children and animals might be considered as having reasons for belief even though they *cannot* be blameworthy for not believing what their evidence supports. Thus, (1) does not seem to hold.

I will grant that these cases would pose a problem for the universal claim that everybody capable of acting and believing for reasons can (in some sense) be blameworthy for non-

¹⁸ I explore them by spelling out my account of epistemic and other forms of attitudinal blame in chs. 4.3, 6.4, and chs. 10 and 11.

¹⁹ I am grateful to Matthew Chrisman for pointing this out to me.

compliance with decisive reasons. I will just argue that this is not what (1) states. The objection rests on a misunderstanding of the premise.

I agree that most children of a certain age and many animals can act and believe for reasons – this fact explains why their bodily movements and mental states are different from brute happenings. When an animal or a child walks around, this is not like a leaf being blown around by the wind. Rather, animals as well as children are guided by their interests, desires, beliefs, and sometimes by their deliberations.²⁰ I think it is a tricky question whether children and animals are *never* blameworthy for not doing or believing what is decisively supported by reasons, but I need not get into these issues here.

The objection can easily be met by fixing the reference of ‘we’ in premises (1) and (2). By ‘we,’ I primarily refer to most human adults. Even if children or animals were never blameworthy for failing to properly base their beliefs on their evidence, this would not show that evidence does not provide us with reasons for belief. Rather, for evidence to provide us with reasons for belief it is necessary that most human adults, who possess the relevant factual and moral knowledge as well as the capacities necessary in order to count as fully responsible agents, can be blameworthy for not properly basing their beliefs on their evidence. There might be cases of children and animals who can be blameworthy insofar as they possess the relevant preconditions on being a responsible agent. These cases, however, do not pose any problem for (1), because (1) does not talk about these cases. And the cases which seem to pose a problem – animals and children which are blameless and yet act for reasons – can be excluded merely by fixing the reference of ‘we’ in the way just described.

4.2.2 A WORRY FROM OBJECTIVISM ABOUT ‘OUGHT’ AND ‘REASONS’

Premise (1) is not controversial for objectivists about the meaning of ‘ought’ and ‘reasons.’ Objectivists deny a close connection between failing to do what one ought to and being blameworthy or praiseworthy (a connection usually utilized or argued for by subjectivists).²¹ Objectivism states that ‘S ought to ϕ ’ means that ϕ ing is the best option, no matter whether S is in a position to know or has some kind of cognitive access to whether ϕ ing is the best option. For example, if my house is burning even though I do not have any clue that it is burning yet, the

²⁰ It is important to note, however, that the fact that the movement of an animal is not like a movement of a leaf which is blown around by the wind is not sufficient to show that the animal acts for a reason. Arguably, worms are not like leaves, but they do not act for reasons. Acting for reasons requires, at least, having certain attitudes. It is doubtful that worms have attitudes in a non-metaphorical sense.

²¹ For some discussion and an argument for subjectivism that builds on the notion of praiseworthiness, cf. Lord 2018, ch. 8. See also Kiesewetter 2017, ch. 8 for a good overview and another case for subjectivism.

objectivist would claim that I ought to leave the house. The subjectivist would deny this and say that I only *ought* to leave the house if I – in some way or other – have cognitive access to the fact that the house is burning. If I have access to that fact but do not leave the house, then I am blameworthy. According to the objectivist, it could be the case that I am not blameworthy, and yet it is true that I *ought* to leave the house (namely, in the case in which I am not in a position to know that it is burning). It thus seems that, if we are objectivists, we do not think that in order for it to be sometimes true that we ought to ϕ , or have reasons to ϕ , it must be possible for us to be blameworthy *merely* for not- ϕ ing – for according to the objectivist, I am never blameworthy merely for failing to do what I ought to do, but only in combination with the fact that I have cognitive access to the fact that I ought to ϕ .

However, (1) is uncontroversial for objectivists because it states merely a *loose* connection between reasons and blameworthiness. Even if we grant to the objectivist that we can be completely ignorant of our reasons, there will at least be cases *conceivable* in which we are blameworthy for failing to give the response that is best (or correct). When we focus on actions rather than attitudes, such cases will be cases where we either act against our better knowledge of what is best or where we culpably fail to know what is best and do the wrong thing as a result of our ignorance. Analogously, if we are objectivists about reasons for belief (i.e., we think that we should always believe what is true rather than what is supported by our evidence), then we can still argue that we are at least *sometimes* blameworthy because we fail to have a true belief – namely, in at least some of the cases where the evidence was accessible to us, and yet did not believe what it supported.

We can read the notion of ‘*our* evidence’ employed in the argument as an in-built accessibility condition. There might be some room for dispute whether there is evidence that is not accessible to anyone – e.g., some dinosaur bones buried deep down in the ocean. This will crucially depend also on ontological questions about what kind of thing evidence is: is the dinosaur bone evidence, or are our *appearances* of the dinosaur bone evidence, or is evidence an abstract object, like a fact, i.e., a Fregean thought? It seems that appearances might be necessarily accessible to the subject, while physical objects or facts need not. No matter how we answer this ontological question, when I talk about ‘our evidence,’ I mean the evidence that is accessible to the person. Premise (1) of the argument from blameworthiness is thus not affected by accepting objectivism about reasons for belief: even if we always ought to believe what is true, we can grant that this ‘ought’ has normative force only if we can be blameworthy for non-compliance with this ‘ought’ (i.e., in those cases when what is true is accessible to us).

4.2.3 A WORRY FROM EPISTEMIC INSTRUMENTALISM

Steglich-Petersen (2011) might resist (1) by pointing out that, according to his analysis of ‘ought to believe,’ there *are* reasons for belief even though he accepts that we are not blameworthy merely for not believing what one’s evidence (to which one attends) sufficiently supports. Take, again, his analysis of reasons for belief: ‘Necessarily, if S has all-things-considered reason to form a belief about *p*, then [if S has epistemic reason to believe that *p*, S ought to believe that *p*]’ (2011, 24). This analysis implies that there are reasons to believe that *p* if two types of constituent reasons – a practical and an epistemic one – are present. Thus, Steglich-Petersen can grant that we are never blameworthy merely for failing to properly base our beliefs on our evidence – i.e., for not believing what our *epistemic reasons* sufficiently support. But we can still be blameworthy for not believing what our *reasons for belief* support, because these are not merely epistemic reasons, but rather epistemic reasons plus a practical reason.

However, by doing this, Steglich-Petersen just switches the topic, and leaves the normative force of epistemic reasons in the dark. As I pointed out in more detail in subchapter 4.1.2, the ‘ought to believe’ Steglich-Petersen’s analysis is *conditional* on reasons for action (i.e., on ‘all-things-considered reasons to form a belief about *p*’). The violation of *unconditional* epistemic norms, like (ER) and (ER*), does *not* imply blameworthiness according to Steglich-Petersen. Thus, there are no unconditional reasons for belief for Steglich-Petersen, and thus no reasons for belief in my sense of the term ‘reasons for belief.’ Steglich-Petersen’s strategy of making the ‘ought to believe’ conditional on reasons for action does not preserve the normative force of object-given reasons for belief. Rather, it just grants that object-given reasons for belief (i.e., *epistemic* reasons) do not have any normative force on their own. That is why Steglich-Petersen takes refuge in a pragmatic version of epistemic instrumentalism.²²

²² There are non-pragmatic forms of epistemic instrumentalism. An example for a non-pragmatic form of epistemic instrumentalism is epistemic teleology: we ought to base our beliefs properly on our evidence because we thereby make it likely to reach the aim of having a true belief. On the aim of belief, cf. esp. the contributions in Chan 2013. Another version of epistemic instrumentalism holds that by being epistemically rational, we likely maximize the epistemic aim of holding true beliefs. These positions provide an extensionally inadequate account of epistemic rationality. For according to these positions, it would be rational to adopt a belief that is insufficiently supported by the evidence if one thereby maximizes one’s true beliefs. Yet one’s belief does not cease to be irrational just because it produces more beliefs in the future that are true or rational. Rather, a belief’s rationality is evaluated independently of any beliefs one does not yet hold.

4.2.4 A WORRY FROM EPISTEMIC PERMISSIVISM

One might want to object to (1) by adopting a permissivist epistemology. Permissivism states, roughly, that our total set of evidence permits more than one set of doxastic attitudes to take towards each (or at least some) proposition(s).²³ According to a permissivist, it could be that when we have sufficient evidence for a proposition, it is both epistemically permissible to believe it as well as epistemically permissible not to believe it. Such an account might seem to be exactly the conclusion to draw from Harman's clutter-objection (cf. 4.1.1): we are not rationally *obligated* to believe anything that our evidence sufficiently supports, and thus not blameworthy. But propositions that are sufficiently supported by our evidence are at least always *permitted* to believe.²⁴

However, it is not straightforward how permissivism could pose a problem for (1). The premise merely states the following necessary condition on reasons for belief: that we can be blameworthy for failing to properly base our beliefs on our evidence. To deny this, the permissivist would have to argue that this is not a necessary condition for reasons for belief. They would have to say that evidence can be said to provide us with reasons for belief because sufficient evidence can make it permissible to draw certain conclusions, and that this permissibility does not imply that one can be blameworthy for failing to properly base one's beliefs on one's evidence.

But this last claim is false. If something is permitted only under a certain condition, then it is *not* permitted – and thus *prohibited* – if this condition is not fulfilled. That is, that I am only permitted to believe that p if p is sufficiently supported by my evidence implies that I *ought not* to believe that p whenever p is not sufficiently supported by my evidence. Thus, even if we understand the normative force of evidence in terms of permission, this will still require us to make sense of the idea that we sometimes *ought* to have certain doxastic stances (i.e., we ought not to believe what is insufficiently supported by our evidence). The phrase 'failing to *properly* base our beliefs on our evidence' in the argument from doxastic blameworthiness can then just be understood as equivalent to 'failing to withhold our belief in propositions that are not sufficiently supported by our evidence.' That we ought to withhold such beliefs is what

²³ The denial of permissivism is often discussed as the *Uniqueness Thesis* (as introduced by Feldman 2007). On epistemic permissivism and some of its problems, cf. White 2005.

²⁴ Thanks to Conor McHugh for pointing out the option of permissivism to me. Asbjørn Steglich-Petersen agreed that he is not ready to commit to the claim that one is not always allowed to believe what one's sufficient evidence supports. He merely claims that one is not always *blameworthy* if one does not believe upon sufficient evidence, and that one is not always required to believe upon sufficient evidence. (This exchange took place at the *Ethics of Mind* conference in Erlangen in July/August 2017.)

(ER*) claims. I have already argued in chapter 4.1 that (ER*) seems to be as implausible as (ER).

Thus, if one reformulates (ER) as a claim about permission rather than an obligation, one commits to (ER*), and thus does not make the normative force of reasons for belief any more intelligible. If we cannot be blameworthy or criticizable merely for believing what is insufficiently supported by our evidence, then it is both mysterious why we *ought* not to believe such insufficiently supported propositions, and, consequently, why we are *allowed* to draw implications from our beliefs only when these implications are sufficiently supported, but not otherwise.

Furthermore, we should resist understanding the normative force of evidence in terms of permissibility. For that we are sometimes permitted to believe what our sufficient evidence supports (but not otherwise) does not show that evidence provides us with reasons. The reason why we are usually permitted, but not obligated, to believe what is sufficiently supported by our evidence does not stem from the fact that evidence provides us with reasons. There is another, more plausible, explanation: usually, our rational capacities just function the way they do, and we cannot change this. It is this inability we often experience when we believe that explains why it is usually permissible to believe what is sufficiently supported by our evidence. The explanation is not that evidence provides us with reasons. I will support this explanation of the phenomenon of epistemic permissibility in the remainder of this subsection.

In order to support the idea that it is permissible to believe that *p* whenever we have sufficient evidence for *p*, one might give the following example. Imagine that you have sufficient evidence for *p* and *p* describes something absolutely trivial, like the number of dust specks on your table. However, in this case it is not quite trivial whether you *form* the belief that *p*. If you form the belief, you will cause the cruel death of some person by some complex causal chain, and you know of this. Nevertheless, one might argue, it would be inappropriate to blame or criticize you for believing that *p*, and thus for causing a cruel death. This is because you have so much evidence for the belief that *p*. After all, what can you do about whether you form the belief?

I agree that you might be excused in this case, and that it might thus be said that it is permissible for you to have the belief. But this is not because evidence provides you with reasons for belief. Rather, this is due to your *inability* to avoid forming the belief. If you were able to avoid forming the belief, for example by trying not to attend to your evidence in order to forget about it, or by knocking your head against a wall, then you would be blameworthy for ending up with the belief if you did not try to avoid it by such means. Thus, the fact that usually

you are permitted to believe what is supported by your evidence (even if it has bad consequences) does not show that evidence is or provides you with reasons. It just shows that, as rational beings, we often cannot resist believing what our evidence supports.

But isn't there an analogous phenomenon in the case of practical reason that does not allow us to say that the permissibility in question does not arise from reasons? If we had a justified but yet false belief that we have decisive reasons for an action, then acting on this belief might make it the case that we were, in some sense, unable to do otherwise, and thus our action would have been permissible. Furthermore, it would have been permissible *because* we suffered this normative inability. Analogously, our inability in the domain of belief might be normative as well, and not merely forced on us like an inability to move when one is bound to a chair.

Yet the inability we sometimes experience in the face of practical reasons is different from the one we experience when we have a great deal of evidence. It is not the case that, literally, Martin Luther could not do otherwise when he said 'Here I stand. I can do no other' – rather, his incapacity 'is at once a genuine incapacity and yet in no way compromises [his] agency or self-control' (Watson 2002, 100). Luther was not *physically* incapable of doing otherwise, as someone who is tied to a chair is physically incapable of getting up. Rather, the incapability was in *normative*: given that he justifiably and non-culpably believes that this is the right thing to do, he can do no other.²⁵ Even if it would turn out against the odds that what he did was wrong (assuming objectivism about reasons), his action would still be blameless, and probably even praiseworthy.

The incapacity of someone who believes what their evidence supports, by contrast, is more akin to physical incapacity. Insofar as our rational belief-forming dispositions are functioning in a normal way, we can do no other than believing what our evidence supports. Due to this physical incapacity in a given situation, we might be *excused* for believing something that is supported by our evidence, even if believing has bad consequences. But it is not the case that, *if we had even more evidence*, it would be the case that we *ought* to believe what is supported by it. There is no point at which our believer, whose belief results in a horrible death, could, similar to Luther, exclaim with this kind of rational self-determination 'Here I stand. I can believe no other.' Given that their belief results in a horrible death, this statement would be utterly inappropriate. Doxastic incapacity is not comparable to Luther's volitional necessity, but rather seems to be much more brute, or physical. In the case of practical reasons, more

²⁵ For a detailed discussion of the sense in which volitional necessity is 'normative,' see Watson 2002, 102-122.

practical reasons can create an obligation out of what was merely permissible before those reasons were present. If the reasons are weighty enough, this may even cause volitional necessity. Yet more evidence neither transforms permissible beliefs into obligated ones, nor into ones we experience as necessitated by our normative reasons. This, however, is what we would expect if evidence would provide us with reasons to believe in the way considerations about value or goodness provide us with reasons to act.

Thus, the notion of permissibility does not help us with understanding the normative force of evidence. By introducing permissibility in our formulations of epistemic standards, we thereby do not eliminate blameworthiness as a necessary condition on reasons for belief as stated in (1). Furthermore, the reason why it is usually permissible to believe what one's evidence supports does not stem from the fact that evidence provides us with reasons. We might say: epistemic permissibility is not a normative category at all. We thus should think about the way we blame and criticize each other for epistemic failures in order to understand the normative force of evidence, rather than about how evidence can make certain beliefs permissible.

4.3 SOME PERSPECTIVE: THE WEAKNESS OF THE ARGUMENT FROM DOXASTIC BLAMEWORTHINESS

The purpose of the present part of this book is to spell out the threat from mental nihilism as forcefully as possible. The outlines of a reply to the challenge will be provided especially in part IV. Yet I should already say where I think the argument from doxastic blameworthiness goes wrong so as not to leave the reader in the dark. I will thus provide some perspective of how I think we should ultimately respond to the argument from doxastic blameworthiness.

I have argued that (1) states a very loose connection between reasons for belief and blameworthiness, and that (1) is not called into question by the *prima facie* worries I discussed above. The reason for this plausibility was also that I employed a loose use of the word 'blame,' according to which each kind of criticism of a person that somehow presupposes their responsibility would count as blame. It is important to see that (1) would be false if we assume that the relevant sense of 'blame' is a passionate form of *moral* blame. For sure, a person's irrationality does not, by itself, usually give rise to hostile emotions like resentment, indignation, or guilt. If someone believes that candidate X will win the next elections because the flight of the birds gave a sign, then our reactions to that person seem to be quite different.

Most importantly, I think there are reactions to such cases which we should count as a form of criticism which is only appropriate towards responsible beings. Yet these reactions do not presuppose any form of control, but rather that we stand in specific relationships with one

another. I will thus deny (2). I will later argue that we can become aware of the forms of criticism I have in mind when we think about how we blame or criticize each other for attitudes other than beliefs: certain emotions that define our interpersonal relationships as well as the attitude of disrespect will play important roles in my outlines of an account of mental responsibility in part IV. I will not provide a detailed account of epistemic blame in part IV, however. This is because I do not think a detailed account of each kind of mental blame or criticism is necessary for providing the outlines of a solution to the problem of mental responsibility.

Yet I should point out what I think are promising candidates for epistemic blame. Kauppinen (2018) proposes that epistemic blame is a form of *distrust*. To distrust someone epistemically is, in some way, *more* than to just not treating them as a reliable source of information – as we would do with an unreliable computer that unreliably predicts us certain outcomes.

We might distrust Sarah because we suspect that Sarah is a spy. In that case, we think that she is pretending and lying, and thus hiding some important information about herself from us. Although we might see this as a paradigm case of distrust, I do not think that it is necessary for distrust that the person is hiding some important information about themselves from us. To see this, consider what I would view as a paradigm case of *epistemic* distrust. I might not trust someone insofar as I am not ready to provide them with an information because I fear that the person will draw the false conclusions from the information. Take Tom:

Tom the Paranoid

Tom always irrationally suspects other people to conspire against him. If I tell Tom that two of his colleagues recently meet more often for lunch than usual, he might interpret this – irrationally – as evidence that they conspire against him. Tom does not hide anything from me. Rather, he is quite open about his fears. I rather feel as if he is hiding something from *himself*: maybe a fact about his own past that would explain why he draws such conclusions hastily.

Tom is not trustworthy epistemically because he is in inner conflict. His epistemic irrationality is a manifestation of some vice – a defect in his character. Distrusting Tom is compatible with assuming that he could not have avoided the rational defect he suffers. Yet my distrust is grounded in this defect, and *only* in this defect. I do not want to provide Tom with certain information because I do not trust him that he will understand its implications. Depending on the degree of his irrationality, I might even doubt that he has the beliefs he thinks he has.

Even though legitimate distrust seems to presuppose that one's epistemic irrationality is not just an occasional lapse, there is *nothing more* I need to know about Tom than that he

(regularly) fails to properly base his beliefs on his evidence when it comes to certain topics in order to legitimately distrust him epistemically. I can view my epistemic relationship to Tom as impaired in certain respects – for part of this relationship is that we regard each other as trustworthy sources of information, and that we know about each other that we draw the right conclusions from the information we provide each other with. We cannot view our relationship with an unreliable computer as impaired in this way, because we do not have relationships with computers, and because they do not display any vices or could hide things from themselves in a way a person can.

Let me point out why I think Tom is a plausible counterexample to (2). I think it is correct that we are not *necessarily* blameworthy or criticizable for failing to believe what our evidence supports (even if we attend to it). A person is not blameworthy in any sense of the word for occasional lapses in their epistemic rationality – as people are also not blameworthy for occasional lapses in their moral conduct (say, you accidentally forget to call me even though we said to talk to each other today, but only because you were very stressed). However, as soon as there is some kind of *pattern* behind these cases of irrationality – if irrationality cannot be excused as an occasional lapse –, we view our relationship to irrational people as impaired in specific ways, e.g., by partly losing our epistemic trust. If irrationality adds up in such a pattern, it begins to *matter* for our relationship to the irrational person in a straightforward way (the relationship gets worse as a relationship), and consequently for our actions related to the person. Yet our blame is not grounded in the fact that the person violated any *practical* norm. It is *merely* grounded in the fact that the person violated (and can be expected to violate again) the standards of rationality.

One might worry at this point that *Tom the Paranoid* is not a case of purely epistemic blame insofar as practical stakes are involved. My reply is that, even if there are practical stakes involved, and even if this is necessary for someone to be epistemically blameworthy, our blame is not grounded in these practical stakes, but rather in epistemic irrationality. We can see this by the fact that our intuition that it is appropriate to distrust Tom does not fade even if we assume that his epistemic irrationality is not itself culpable – i.e., even if it did not result from a practical failure of attitudinal management. In such a case, the only norm-violation our distrust can be grounded in is his epistemic irrationality.

This is meant to be an appetizer for my account of responsibility for attitudes which I spell out in more detail in part IV. However, I think that we need to understand our challenge more clearly before we can present a fully satisfying reply to the challenge. The rest of part II is devoted to this task (except for chapter 6.4). In the remainder of the present chapter, I would

like to point out the limits of *The easy solution* which we can now already see after having the argument from doxastic blameworthiness as a challenge in front of us.

4.4 CONSEQUENTIAL CONCEPTIONS OF DOXASTIC RESPONSIBILITY

We cannot reply to the argument from doxastic blameworthiness by pointing out that in believing we sometimes violate *non-epistemic* norms, and that this is what makes us blameworthy. Pointing this out would not be in conflict with premise (2). Yet some recent accounts in the ethics of belief point to the violation of requirements of actions that *cause* beliefs: they argue that responsibility for belief can be made intelligible by the fact that we have indirect control over belief by means of various ordinary activities.²⁶ According to these philosophers, we never blame people merely for having a belief that is not properly based on our evidence. Rather, we blame them for their belief only if their belief is culpable, i.e., if they failed to manage and influence their doxastic life in the way they should have done: we blame them for their beliefs, for example, because they *earlier* did not attend to certain evidence or did not investigate properly. Responsibility for belief turns out to be completely analogous to our responsibility we have for other consequences of our actions (Meylan 2017). We can fail in our doxastic lives only if we have these lives as a result of not *acting* the way we ought to. Thus, these authors endorse what I have called *The easy solution* to the problem of mental responsibility with respect to belief (cf. ch. 0.1).

However, these accounts cannot explain what I call *purely* epistemic responsibility. Purely epistemic norms are supposed to be standards governing *belief*, not action. They say that our beliefs ought to be properly based on our evidence. That epistemic norms govern beliefs rather than actions can be seen in cases in which we are practically required to cause ourselves to violate an epistemic norm. If I can avoid disaster by failing to properly base a belief on my evidence, then I *should* ensure that I violate an epistemic norm. For failing to be epistemically rational once does not seem to be a huge cost – or a cost at all. Yet such a failure of epistemic rationality is still a *failure* in some sense. Clearly, it is no failure to act the way we ought to have acted. For in the case at issue we did what we *ought* to have done – we *caused* ourselves not to comply with an epistemic norm, and thus caused ourselves to commit a *purely doxastic* failure: a failure in *believing*, not acting. We complied to a practical norm but violated

²⁶ Cf. Chrisman 2008; 2016; Meylan 2013; 2017; Peels 2017. The details of these accounts differ significantly, yet they share the common feature of making doxastic responsibility or epistemic normativity intelligible through indirect doxastic control.

an epistemic one. Such cases show that epistemic norms are independent of practical norms: we cannot explain our responsibility to these norms by pointing to actions prior to belief.²⁷

In response, accounts that explain the normativity of belief through indirect control could just deny that we are ever directly responsible for complying with what I call ‘purely epistemic norms,’ thereby endorsing anti-responsibilism. However, if they deny our responsibility to epistemic norms, then they either deny the very weak connection between blameworthiness and reasons for belief which is central to the argument from doxastic blameworthiness (as pointed out in 4.1.3), or they are committed to doxastic nihilism, i.e., to a denial of the existence of reasons for belief. The latter option is radical, and we should only go for it if we do not find a plausible account of how we are responsible for complying to purely epistemic norms.

If these accounts choose the first option, then they say that in order for there to be reasons for belief, it is not necessary that we can be blameworthy for not complying with epistemic reasons. They thus reject premise (1) of the argument from doxastic blameworthiness. Yet this would leave us puzzled why epistemic reasons are *reasons*: it would leave unexplained why we can regard purely evidential considerations as exerting normative force on us. That is, these accounts would leave the phenomenon of epistemic normativity unexplained.

If we allow for concepts of blame or criticism in terms of impairments of relationships, as I sketched in section 4.3 above, then we open up the space for a face of responsibility that does not presuppose that the one’s we blame or criticize had indirect control over their attitudes. Thus, consequential accounts of doxastic responsibility cannot account for this face of responsibility for beliefs. If we allow for this face of responsibility, then we might be able to make room for something like *purely* epistemic blame: if we are blameworthy *morally* as soon as our relationship to our *moral* community is impaired, we might say that we are *epistemically* blameworthy as soon as our relationship to our *epistemic* community is impaired. This latter impairment might exist as soon as one is epistemically irrational, and this impairment might matter in specific ways for how we should relate to one another.

Yet it is still unclear as to what such epistemic blame exactly amounts to: what are the reactive attitudes, besides epistemic distrust, that we show towards those who are epistemically

²⁷ For this point, see also David Owens (2000, 87), who argues ‘that our control over [...] actions yields only an *indirect* control over belief and[...] such indirect control cannot underwrite the *direct* responsibility of belief to epistemic norms.’ We can obviously violate an epistemic norm without violating any norm of action. Thus, epistemic norms cannot be made intelligible by reference to indirect control, i.e., belief-control through action. I discuss cases in which our practical requirements conflict with our epistemic norms in ch. 6.

irrational *merely* because they are epistemically irrational?²⁸ I do not have a fully satisfying answer to this question. Rather, I explained why spelling out what epistemic blame amounts to is an important project: doing this is necessary, I think, to make epistemic normativity fully intelligible. Some recent accounts that reduce our duties to believe to duties to act tend to ignore this challenge by talking past the phenomenon of (purely) epistemic normativity. Even accounts that do not talk past the phenomenon by acknowledging the existence of purely epistemic norms leave our responsibility to these norms unexplained if they do not spell out as to what purely epistemic blame amounts to.

4.5 SUMMARY AND OUTLOOK

I will *not* meet *this* challenge in the present book: fully understanding the normative force of evidence. Rather, I want to provide a solution to the problem of mental responsibility (cf. chapter 2): How can we be directly responsible for our attitudes even though we lack any form of control over our attitudes that could explain our responsibility? I have argued in chapter 2 that the direct responsibility in question must arise from the fact that our attitudes are subject to the standards of attitudinal rationality, and thus that the problem of mental responsibility requires us to think about the status of the standards of rationality. The present chapter spelled out a challenge for the very idea that there are standards of epistemic rationality by calling into doubt the status of evidence as providing us with reasons to believe (4.1-4.2): if we are never blameworthy merely for properly basing our beliefs on our evidence, and if a minimal connection between blameworthiness and reasons holds, then there are no (epistemic) reasons for belief. I also sketched my reply to this challenge (4.3), and I pointed out the limits of *The easy solution* by considering consequential conceptions of doxastic responsibility: these conceptions, and thus *The easy solution*, cannot account for our responsibility to epistemic norms (4.4). Solving the problem of mental responsibility will ultimately require me to sketch a general account of the normative force of epistemic reasons and other object-given reasons for attitudes, but the details of an account of the nature of epistemic blame will be left open for another investigation.

I now continue spelling out my challenge of solving the problem of mental responsibility by first extending the threat from doxastic nihilism to a threat from mental nihilism (chapter 5), and then reinforcing it with another argument, to which I provide the outlines of a reply

²⁸ A reply was recently provided by Brown (2018), who defends the idea that we are blameworthy epistemically in a broader sense of 'blameworthy' than we usually have in mind when thinking about blame (her response inspired my account in part IV), and Kaupinnen (2018), who argues that epistemic blame is epistemic *distrust*.

(chapter 6). In chapter 6.3 as well as in the ‘Summary and Conclusion of part II’ after chapter 6, I will provide the framework of my solution to the problem of mental responsibility. I spell out the details of this solution to my challenge in part IV (after having considered the pragmatist’s reply to the challenge in part III) by giving the outlines of an account of the normative force of the standards of rationality in terms of responsibility and blame. These outlines suffice for an explanation of how mental responsibility is possible even in the absence of any special form of direct mental control. Providing this explanation is the challenge that I will meet in this book.

Chapter 5

The Argument from Mental Blameworthiness

In this chapter, I present a generalized version of the argument from doxastic blameworthiness (5.1). This generalized argument concludes that there are no reasons for attitudes whatsoever: the argument from *mental* blameworthiness. I argue that, maybe surprisingly, the argument does not lose any of its *prima facie* plausibility when we apply it to desires and intentions. Yet when we apply it to certain emotions, the argument does not seem as plausible anymore.

I conclude that, even though we see a weak point of the argument when we apply it to certain emotions, the argument still presents us with a challenge of understanding the normative force of object-given reasons for attitudes (5.2). I argue that this poses a new challenge to recent defenses of the normativity of rationality – a challenge which has gone largely unnoticed. I finally specify the challenge by distinguishing different concepts of normativity (5.3). The next chapter then provides more substance to this challenge by presenting another argument to the effect that there are no reasons for attitudes, and allows us a first step towards understanding mental normativity by replying to this argument.

I will turn to a reply to this challenge in part IV of this book. But fully meeting it will be a task beyond the scope of the present inquiry. The focus of the present book is on solving the problem of mental responsibility by evaluating the plausibility of *The easy solution*, which only requires the outlines of an account of the normative force of object-given reasons that allows us to see how we can be directly responsible for non-compliance with these reasons. A full reply to the threat from mental nihilism would include a detailed account of the different forms of blame that are appropriately directed at different attitudes in virtue of the fact that someone violates the standards of rationality of that attitude.

5.1 OBJECT-GIVEN REASONS AND BLAME

I first present the generalized version of the argument from blameworthiness: the argument from *mental* blameworthiness (5.1.1). I give some initial clarifications. I then focus on the controversial premise (2) of the argument (5.1.2). I first outline which kinds of cases would support (2) and why. I then spell out and discuss such cases for desires (5.1.3), intentions (5.1.4), and emotions (5.1.5). My discussion of the argument from mental blameworthiness will show that it loses some of its persuasiveness when we look at emotions, but that the

argument from mental blameworthiness seems (maybe surprisingly) *as persuasive as* the argument from doxastic blameworthiness when we focus on desires and intentions.¹ I summarize these results (5.1.6).

5.1.1 GENERALIZING THE ARGUMENT FROM BLAMEWORTHINESS

Here is the generalized version:

The argument from mental blameworthiness

- (1) Considerations of kind K are reasons for or against an attitude only if we can be blameworthy just because we do not have an attitude that seems to be decisively supported by considerations of kind K.
- (2) We cannot be blameworthy just because we do not have an attitude that seems to be decisively supported by object-given considerations.
- (3) Thus, object-given considerations are, it turns out, no reasons for or attitudes.

If we assume that only object-given considerations could possibly be reasons for attitudes (as I do here),² then it follows from (3) that there are no reasons for attitudes whatsoever. This is an unsettling conclusion which we should not accept without further ado. Yet before we look where our arguments from blameworthiness go wrong, we should first consider whether the generalized version of the argument is as plausible as the argument from *doxastic* blameworthiness from chapter 4. I will first argue in this subchapter that the argument from *mental* blameworthiness is persuasive to the same degree as the argument from *doxastic* blameworthiness when we focus on desires and intentions. Yet we will see that the argument loses some of its persuasiveness when we focus on specific emotions.

Some clarifications first. The argument uses the phrase ‘seems to be supported by considerations.’ This reflects the ungrateful position the mental nihilist is in: they have to argue that what other people call ‘reasons’ for attitudes are in fact no such things – object-given

¹ To be precise: I do not think that any of the arguments is ultimately persuasive. Still, they pressure us to point out where they go wrong, because they have intuitive plausibility.

² For my distinction between object-given reasons and state-given reasons, cf. ch. 0.2.4. I argue against practical reasons for belief throughout part III of this book. There might be some state-given considerations that are not practical (e.g., the consideration that it would be rational to have an attitude, cf. Kolodny 2005, 551). However, it would be odd if there were state-given reasons for attitudes but not practical reasons. If practical considerations (e.g., about the pleasure of having a belief) fail to be reasons for belief, then this is because they in general of the wrong kind for counting as a reason: they are state-given. I will assume throughout this book that state-given reasons stand and fall together with practical reasons, although the latter are merely a subclass of the former.

considerations do not in fact *support* an attitude. It merely *seems* as if they do. There might be various ways in which this *seeming* might be spelled out – as merely referring to the epistemic position of others (most people and philosophers tend to assume that object-given considerations *in fact* support attitudes) or rather as grounded in some *other* relation that in fact holds between attitudes and object-given considerations, even if this relation is not a relation of support (but can easily be confused with it). A version of the latter strategy would hold that object-given considerations seem to support attitudes because it matters *in most cases* whether we hit the truth, desire only what is desirable, and intend only what is right. That is, the relation between object-given considerations and the attitudes they seem to support is in fact one of making it likely that it matters whether we have a specific true belief, desire that is directed at the desirable, or right intention.

I need not go into the details of spelling out the ‘seeming support’-relation here because I will not commit myself to nihilism. Spelling out this notion would be the task for the nihilists themselves. According to my position, object-given considerations are in fact reasons that support attitudes. I do not think that the fact that nihilists have to say something about the relation between object-given considerations and attitudes without making this relation a support-relation is a decisive objection against their view. There might be some plausible story they can tell about why the relation *seems* like a relation of support along the lines of the strategies mentioned above.

Let me comment on the intuitive plausibility of the premises now. Premise (1) is plausible for the very same reason as the first premise of the argument from *doxastic* blameworthiness: it is a very minimal claim about how reasons and blameworthiness are connected. It must be *possible* to be blameworthy for non-compliance with these reasons. Note, as I have argued already in the context of the argument from doxastic blameworthiness above (cf. 4.2.2), that this premise should not be controversial for objectivists about the meaning of ‘ought’ and ‘reasons’: even if objectivism is true, we would sometimes be blameworthy *when we know what attitudes we ought to have* but fail to have them nevertheless.

Let us thus turn to premise (2) and see whether it is as plausible as the second premise of the argument from doxastic blameworthiness.

5.1.2 HOW TO SUPPORT THE SECOND PREMISE: REASONS FOR ATTITUDES AND BLAMELESSNESS

The second premise of the argument from doxastic blameworthiness was supported with types of cases of which the dust-speck case was an instance: S has sufficient evidence for a trivial

proposition, fails to believe in this proposition, and is yet not blameworthy. Such cases support (2), I have argued, because if evidence would provide us with reasons, we would expect that we are blameworthy at least in *some* such cases where all practical considerations are excluded. For if we change the structure of the case by bringing in practical considerations (e.g., cases of racist and sexist beliefs or cases of disbelief in human-induced climate change), then we can no longer guarantee that the blame in question is *epistemic*. Analogously, in order to identify a purely *mental* blame – i.e., blame that is directed at people for their attitudes for not complying with their reasons for these attitudes –, we need to find cases in which all practical considerations – i.e., considerations about an attitude’s value – have been isolated and only the object-given considerations are left. If we are never blameworthy in such cases, it seems that we made a good case for (2). Thus, in support of (2), we are looking for cases in which

- (a) we would usually (assuming that we are not nihilists) judge that the attitude is decisively supported by object-given reasons and
- (b) no practical considerations are at play, like the high value of the attitude itself or its consequences for other people, and
- (c) the subject is not blameworthy.

I will discuss such cases in turn. Note that just finding one such case is not sufficient for establishing (2). Rather, for (2) to be plausible, it needs to be plausible that *in all structurally analogous cases* the subject is not blameworthy as well.

5.1.3 REASONS FOR DESIRE AND BLAMELESSNESS

Consider, first, desires.³ We would expect to find a case where desires seem to be decisively supported by reasons and yet we are not blameworthy for failing to have the desire by focusing on desires with trivial content, as I did when I discussed reasons for belief. However, my trivial desire to scratch my neck because it is itching is not clearly a desire that does not matter. After all, if my desire seems to be decisively supported by object-given reasons, then this is because, say, my neck is in fact itching. Thus, my trivial desire might have the good consequence that I scratch my neck, and one might thus argue that it indeed matters in some way whether I have the desire. At least in cases in which I have some indirect control over whether I develop the

³ I use a very broad, philosophical, concept of ‘desire’ here, where desiring involves that you rationally ought to think that your desired object is good in some respect. It is the concept of a desire as a ‘pro-attitude’ as it is used by Davidson (1963).

desire or not, our intuitions about my blameworthiness or blamelessness for (not) having the desire could be influenced by my failure to comply with my reasons for action rather than my failure to comply with my reasons for desire. We thus need a case where I clearly have no reasons for bringing the desire about.⁴

There are better cases that seem to fit the bill (a)-(c) above. Suppose that X is in fact desirable and your evidence correctly reveals this to you, but what you desire is impossible for you to achieve. This would be a case, I take it, where you are supposed to have decisive object-given reasons for desiring X by any traditional theory of reasons. Suppose, for example, that you have sufficient evidence that it is desirable to walk on Pluto: you believe that you could enjoy an awesome otherworldly landscape when doing so. Yet, for some reason, you do not desire to walk on Pluto. Are you blameworthy for lacking this desire? After all, you *cannot* walk on Pluto. What, then, is the point of desiring it? Not desiring it might rob you of the pleasure of imagining how nice it is to walk there. But we might well stipulate that you have better things to do than engaging in such imaginative projects. It is unclear in such cases, where your desire does not have any benefit for you or for others, why anyone should regard you as blameworthy.

In such cases, it seems that whether you desire X or not does not matter at all. It thus seems that, *merely* failing to have a desire that seems to be decisively supported by object-given reasons does not make you blameworthy or criticizable. Rather, it seems that it only matters as soon as some practical value that would be realized if you had the desire is at stake. Given premise (1) of the argument from mental blameworthiness (with ‘desire’ inserted for ‘attitude’), it follows from your blamelessness in all structurally analogous cases that object-given reasons for desire are, after all, no reasons. The argument from mental blameworthiness does not lose any of its persuasiveness when applied to desires (compared to the argument from doxastic blameworthiness).

5.1.4 REASONS FOR INTENTION AND BLAMELESSNESS

Next, consider intentions. In wanting to describe a case that fits the bill (a)-(c), we face a similar problem as we did with desires when we focus on intentions with trivial content. The intention

⁴ We could focus on cases where I do not have any means to bring the desire about. However, these cases seem too controversial. First, it is not clear whether a desire that is not controllable by any indirect means is a full-blooded desire that is responsive to reasons. Secondly, even if it was, our intuitions about whether the person is blameworthy for not having the desire might just depend on our (pre-)theoretical intuitions about whether control is necessary for blameworthiness. I thus deem it more helpful to focus on another case where no practical value is at stake at all – as we did when considering the argument from doxastic blameworthiness.

to brush your teeth this morning obviously mattered in some way: without it, you would not have brushed your teeth. Similarly, my intention to scratch my neck because it is itching matters to some degree, and I might thus be plausibly conceived as being criticizable (to a certain degree) for not having it because I failed to realize some value. This criticizability might thus not be grounded in our failure to correctly respond to object-given reasons, but rather in our failure to realize value.

Instead, consider an action which you ought to perform *in the future*. Suppose that attending a conference *in a year* would be the right action for you (cf. Kieseewetter 2017, 190-192). The reasons for attending might include, for example, the opportunity for rich academic exchange, for presenting your ideas, and for making important contacts. It thus seems rational for you to intend *now* to attend the conference *in a year*. Yet there is *nothing lost* if you do not yet intend it. You might be *akratic* right now: you know you should attend, but you do not intend to attend. But since it is still a year until the conference takes place, and thus another year until it *matters* whether you intend to attend the conference, you have plenty of time to overcome this akrasia. Since it does not yet matter whether you attend the conference, you are not blameworthy for being akratic. That is, you are not blameworthy for not intending what you ought to do – i.e., for not having an intention that seems to be decisively supported by object-given reasons. If this is true in all structurally analogous cases, then it follows from your blamelessness in these cases and premise (1) that object-given considerations for intentions are no *reasons* for this intention.

Is it plausible that this case fits (a)? Is it plausible to say that it seems that you have decisive object-given reasons for intending to attend the conference in a year? Kieseewetter does not think that your reasons are decisive in this case. He argues that you do not *yet* have decisive object-given reasons to intend to attend the conference in the future. Rather, it *becomes* decisively supported by object-given reasons as soon you need to intend to attend in order to ensure that you will attend.

However, it is unclear what is supposed to explain the fact that the intention becomes decisively supported by object-given reasons as soon as it matters whether you intend to attend. Let us assume that, as the conference draws nearer, and you need to take steps in order to ensure that you attend, nothing relevant to the deontic status of your *attending* has changed. No further object-given reason to intend to attend has appeared on the horizon. Nevertheless, you now have, according to Kieseewetter, decisive object-given reasons to intend to attend the conference, and thus you would be blameworthy if you fail to intend this now that it matters. But how can object-given reasons can *become* decisive while remaining the same set of reasons?

Kiesewetter's account amounts to saying that object-given considerations for intention gain their normative force only in cases where it matters whether we comply with them. This amounts to granting that these considerations have no such force *on their own*.

Like Steglich-Petersen's background condition on epistemic standards that one ought to believe that *p* only if one has a reason for an action by which one would cause the belief that *p* (the 'all-things-considered reason to form a belief'),⁵ Kiesewetter here introduces a background condition on the enkratic norm that one ought to intend what one believes that one ought to do only if one has a reason to bring it about that one so intends (i.e., a reason one has as soon as it matters whether one so intends). Without such a reason to act as a background condition, it would not matter, and it thus seems that we would not be blameworthy or criticizable for not responding correctly to what seem to be decisive reasons for intention. I thus conclude, *pace* Kiesewetter, that the future-intention case fulfills not only conditions (b) and (c), but also (a).

It is hard to see how we should be blameworthy in structurally analogous cases. I thus conclude that the argument from mental blameworthiness is, when applied to intentions, as plausible as the argument from doxastic blameworthiness.

5.1.5 REASONS FOR EMOTION AND BLAMELESSNESS

Next, consider emotions. The fact that I am confronted with something dangerous would be a paradigm case of a decisive object-given reason for fear (if there were such reasons). But it can seem that this fact is no reason for fear. Imagine you are confronted with a dangerous tiger. However, you are hopelessly bound to a chair when the tiger approaches you, so that fearing it would not help you to escape. You have overwhelming evidence that the tiger is *dangerous*, and thus decisive object-given reasons for fear. However, you would not be blameworthy at all if – for any reason whatsoever – you do not fear the tiger. After all, whether you fear it or not will not be of any help, and you will wrong no one in not experiencing fear.

It seems that whenever we think of a case where (a) you are confronted with something that is dangerous, and (b) it does not matter in any way whether you fear the object, then you are also (c) not blameworthy. We thus have supported premise (2) for the emotion of fear. It follows, given the uncontroversial premise (1), that there are no reasons for fear.

However, it is not as easy to construct such cases with respect to other emotions. Take *admiration* and *love*. Maybe you are blameworthy in some sense if and because you do not

⁵ Cf. chs. 4.1.1, 4.2.3.

admire features of the world that are *admirable*, or if and because you do not *love* someone who is *worthy of your love*. I think whether you are blameworthy will depend on whether we understand admiration and love as stable dispositions or as occurrent feelings. For we might well imagine a case where I am confronted with something admirable (say, I am standing in front of the ancient pyramids), but where I do not *feel* admiration (say, due to my being stressed out from travelling), and where I am not blameworthy (I am excused due to stress). Yet one might think that I would be blameworthy (or criticizable) in some way if I do not have a disposition to feel admiration for what is admirable under usual circumstances. Similarly, if I do not feel the love for my partner in a specific moment (say, because I am stressed out by them right now), this need not make me blameworthy. Yet if I do not have any disposition to feel love for them, I might plausibly be said to be blameworthy or criticizable for this lack of love for someone who would deserve it.

We here have some potential counterexamples against (2). How might the nihilist respond to these cases? The nihilist might argue that we end up with an account of reasons for attitudes that is *unnecessarily complex* if we, on the one hand, accept that what seem to be decisive object-given reasons for one emotion are in fact no such reasons (as in the case of fear), but, on the other hand, that there are in fact such reasons in other cases (as in the case of admiration and love). Rather than committing to such complexity, we should deny that we are blameworthy for failing to admire the admirable and love whom is worthy of our love *in virtue of the fact* that we fail to respond correctly to object-given reasons for these emotions. Rather, we should give another explanation for the blameworthiness involved in such cases. A nihilist might provide the following explanation: When something is admirable or worthy of our love, this essentially involves *practical considerations* in some way, i.e., considerations that explain why it matters to have these emotions. Our failure of appreciating these considerations, rather than our failure to comply with object-given reasons, explain why we are blameworthy. When we have a disposition to feel admiration for the admirable, and a disposition to feel love towards what is worthy of our love, then we thereby acknowledge certain values, and this *acknowledging* is by itself something that is valuable. The reason why we are blameworthy for failing to have these emotions is thus not that we do not have emotions that are decisively supported by object-given reasons, but rather that we fail to realize some practical value in failing to acknowledge the value of admirable or lovable objects.

This objection by the nihilist is not fully clear as it stands. It alerts us to looking closer at what is meant by ‘practical considerations.’ It is hard to see how the practical value of loving or admiring a lovable or admirable object can explain why a subject who fails to acknowledge

these values with the appropriate attitudes is blameworthy or criticizable. For if we are no pragmatists, then the subject cannot admire or love the object *for the reason* that admiring or loving would be of practical value. The nihilist could reply that what we are blameworthy for is for failing to *cultivate* admiration and love – for cultivating admiration and love is something we can do for practical reasons. Yet we can very well imagine cases where there was no reasonable opportunity to cultivate admiration and love, and where our intuitions about the criticizability of those who fail to acknowledge admirability and lovability do not fade: there is still something substantially wrong with a person who fails to feel the way we would expect a rational person to feel when it comes to emotions like admiration and love. Even in the case of fear, we might think that a regular lack of fear in the face of very dangerous situations is a manifestation of some vice, like recklessness.

My suggestion will later (chapter 6.4) be that by failing to be rational, we often display vices. The inability to acknowledge the admirable and lovable is a defect in character that has specific normative consequences for our relationship with others. Why explore the world with someone who cannot admire its admirable features? Why be in love with someone who cannot appreciate a person that would be worthy of their love? Even though these defects in character do not warrant responses like resentment, indignation, and guilt, they warrant criticism or blame insofar as it is appropriate not to get involved with the people who suffer these defects in certain ways. This blame does not merely consist in *not relying* on these people: rather, it is a blame that marks the impairment of specific relationships which we can only have with other responsible beings, and manifests itself, for example, in forms of distrust.

I will return to these issues in chapter 6.3, and in more detail in part IV of this book. There are emotions and other attitudes where the argument from mental blameworthiness loses its persuasiveness. I will argue later that we are necessarily morally blameworthy for showing *disrespect*.⁶

5.1.6 INTERMEDIATE SUMMARY

I have argued that the argument from mental blameworthiness presented in subchapter 5.1.1 has initial plausibility not only when we apply it to beliefs, but also when we apply it to desires and intentions. Yet when we apply it to certain emotions, we might experience doubts about whether it is as convincing as I presented it. According to the argument, we are never blameworthy *merely* for lacking beliefs, desires, and intentions that are decisively supported by

⁶ I return to disrespect in ch. 11.1.

object-given reasons. It seems that it at least must *matter* whether we comply with decisive object-given reasons before we can be blameworthy for non-compliance. Since object-given reasons for beliefs, desires, and intentions do not indicate whether it matters to have the attitude, it follows that a subject's non-compliance to object-given reasons does not imply the subject's blameworthiness (it implies the subject's blameworthiness only if the non-compliance matters). If we accept the plausible premise (1) of the argument from mental blameworthiness, then we are bound to conclude from the subjects' blamelessness that object-given reasons for beliefs, desires, and intentions are, after all, no reasons. If we do not want to be pragmatists either, i.e., if we do not want to commit ourselves to practical reasons for attitudes, then we are committed to nihilism: there are no reasons for attitudes.⁷ This is an unsettling conclusion. What, if anything, went wrong?

I have already argued in chapter 4.3 that there are cases in which we are, in a sense, blameworthy for non-compliance with epistemic reasons. My discussion of reasons for admiration and for love showed that the nihilist's intuition about the subject's blamelessness is hard to hold up when we focus on these emotions. In being irrational, we sometimes display blameworthy vices. This will, roughly, be my reply to the argument from blameworthiness. I thereby reject (2), but accept (1) as a plausible constraint on what can count as a reason.

In the remainder of this chapter, I will, first, point out the relevance of my argument for the current debate on the normativity of rationality, and thereby clarify the dialectic of the challenge I set up here and the possible replies to it, including my own. Secondly, I will distinguish concepts of normativity and point out that the debate about the normativity of rationality is in need of these distinctions in order to clarify the import of the claim that rationality is (not) normative.

5.2 OUR CHALLENGE: UNDERSTANDING THE NORMATIVE FORCE OF RATIONALITY

Let me briefly repeat some central elements of the philosophical discussion about rationality in order to relate my challenge to the debate. Rationality was most prominently understood as mental coherence.⁸ Kieseewetter calls the norms that are supposed to correspond to this kind of rationality 'requirements of *structural* rationality,' because they are concerned with the relation among, or structure of, our mental states. Structural rationality is supposed to require us, for

⁷ Since I did not argue against pragmatism here, this line of thought can well be used in support of pragmatism. In fact, Rinard (2017) is not too far off from arguing in this manner. I argue against pragmatism in part III.

⁸ A famous proponent of this understanding of rationality is John Broome (2007; 2010; 2013).

example, not to intend what one believes one ought not to do; or not to believe what one believes one has sufficient evidence against. These are norms to avoid certain combinations of attitudes. It has been pointed out at length that, if we want to save the idea that these *structural* rational standards are normative standards (that is, if we want to say that we have always a *normative reason* or *ought* to follow these standards), then we face several problems which seem to be unsolvable.⁹ The argument I spelled out in the present chapter and the argument that I will spell out in chapter 6 pose a new challenge for the normativity of rationality, even if we adopt an understanding of rationality that is more promising than rationality as coherence. Let me explain.

Kiesewetter (2017) argues that a more promising understanding for saving the normativity of rationality is rationality as responding correctly to one's reasons (cf. also ch. 0.2.4). The argument for the normativity of rationality, understood as responding correctly to reasons, seems straightforward: 'If rationality consists in responding correctly to reasons, then rational requirements could be understood as inheriting both their content and their authority from the content and authority of the relevant reasons' (Kiesewetter 2017, 160). That is, according to this conception of (attitudinal) rationality, what rationality requires of us is just what we ought to:

(RO) Rationality requires of us to ϕ if and only if we ought to ϕ .

The standard-objection against this conception of rationality is that it might be impossible for us to know what we ought to, but that we are still rational if we respond to our epistemically available reasons while ignoring things that are unavailable to us (Broome 2007, 253). It seems that a defender of response-rationality might plausibly reject the first assumption of this 'quick objection,' as John Broome calls it: we can always know what we ought to do. What we ought to do is determined by the reasons that are *available* to us.¹⁰ This is in line with the idea that not doing what we ought to do implies blameworthiness or criticism. For if one could have known that what one did was wrong, then one is, at least normally and when one lacks an excuse, blameworthy or personally criticizable.¹¹

⁹ The most notable are boot-strapping (Kiesewetter 2017, ch. 4; Kolodny 2005, 514-42) and the challenge of what the reasons are that rationality provides us with (Kiesewetter 2017, ch. 5; Kolodny 2005, 542-7). Cf. also ch. 0.2.4.

¹⁰ This is, a bit simplified, Kiesewetter's strategy (2017, chs. 7 and 8). For objectivism and subjectivism about the meaning of 'ought' and 'reasons,' cf. ch. 4.2.2 of this book.

¹¹ More precisely, one will either be criticizable for not doing what one thinks one ought to, or for not knowing that one ought to (that is, either for akratic action or for ignorance). On responsibility for ignorance, see Rosen (2004). It is important to note that the 'ability to know' needs to be adequately specified. The fact that we can

I will grant here that we might be able to plausibly defend (RO) against the quick objection by rejecting strict objectivism about reasons.¹² Nevertheless, this move does not avoid the challenge I set out for the conception of rationality as correctly responding to (available)¹³ reasons. This is because the challenge arises from a more fundamental assumption in the debate about the normativity of rationality:

(RC) We always ought to respond correctly to our reasons.

How can anyone reasonably call (RC) into question? It might be argued that (RC) is an obvious analytical truth. To do what you ought to do just means that you give the kind of response that your reasons favour most, and to give this response for those reasons that favour it. To give this response for these reasons is to respond ‘correctly.’ Thus, to do what you ought to do just means that you respond correctly to your reasons. Furthermore, what would responding ‘correctly’ even mean if ‘responding correctly’ is not the respond one *ought* to give?

Note first that, if we spell out how the claim is understood by philosophers like Kiewewetter (2017) and Lord (2018), then it is not trivial anymore:

(RC*) We always ought to respond correctly to our *object-given* reasons.

The argument I spelled out throughout this chapter calls (RC*) into doubt. Furthermore, (RC*) is denied by *pragmatists* about what we ought to believe. For pragmatists, there will be cases where our object-given reasons all favour a specific response but in which it is not true that we ought to give the response. For example, our epistemic reasons might all favour the belief that god does not exist, but since – as we can stipulate – it would be better for us to believe in god than not to believe in god (no matter whether god actually exists), our pragmatic reasons for belief would favour believing in god.¹⁴ All things considered, it might be true, according to the pragmatists, that we ought to believe in god, even though the correct response to our *object-given* reasons (here: the *epistemic* reasons) would be not to believe in god.

More generally, we might call (RC*) into doubt because we doubt the *normative force* of object-given considerations – i.e., of the considerations of which many philosophers currently think that they are the only *reasons* for an attitude, i.e., the only considerations capable

in principle know something that is important to know does not always make us criticizable for failing to know it, for there may still be no *reasonable* way to come to know it.

¹² Note that I do not commit myself to a denial of objectivism. I just grant it for the sake of argument. My argumentative aim in this chapter is to pose a new challenge for the normativity of rationality that is independent of objectivism about reasons.

¹³ In what follows, I will omit the condition of availability for the sake of simplicity.

¹⁴ Cf. Pascal (1670, §233), James (1896). For a discussion of this case, cf. Schmidt (2017).

of favouring an attitudinal response. (RC) might in fact be trivial, but (RC*) is not. It seems that we can doubt the normative force of object-given considerations. I spelled out an argument that calls it into doubt throughout the last two chapters. And I am not alone with such doubts – pragmatists have them as well!¹⁵ Even more, pragmatists *deny* (RC*) based on interesting arguments that play a big role in contemporary epistemology. If object-given reasons would lack normative force, then the norms of rationality could not ‘inherit’ any normative force from these considerations. If we then do not want to commit to pragmatism, then it would become mysterious how the norms of rationality should provide us with reasons at all. As non-pragmatists, we need to make this normative force *intelligible* rather than taking it for granted.

There are various ways an opponent of (RC*) might spell out the notion of a ‘correct’ response to object-given reasons for belief. A moderate pragmatist could argue that it is the response the object-given reasons would favour if there were no practical reasons against believing what the epistemic reasons support. For more radical pragmatists, who deny object-given considerations the status as *reasons*, a correct response to object-given reasons could be the response that would be favoured if it was all-things-considered best to hit the attitude’s constitutive aim, e.g. truth (belief), desirability (desire), or rightness (intention). Nihilists can analogously appeal to such all-things-considered reasons for actions that lead to an attitude hitting its constitutive aim, thereby defining a notion of a ‘correct response’ to object-given considerations. Finally, both pragmatists and nihilists might say that what we call a ‘correct’ response is just the response that *seems* to be correct to evidentialists and mental constitutivists. Thus, there are ways for opponents of (RC*) – pragmatists and nihilists – of spelling out the notion of the correctness of a response, and I will thus not treat this notion as worrisome for their position.

Thus, if (RC*) was false, and if we all agree on (RC), then we would be left with only two options. Either we argue that there are other reasons beside object-given reasons that also determine what a correct response is. We would then have to commit to *state-given* reasons. The most obvious and most interesting candidate of such reasons are practical reasons for belief – i.e., reasons which are supposed to favour a belief by showing the belief to be valuable. Thus, according to this first option, it is most plausible to commit to *pragmatism*.

¹⁵ The more ‘radical’ pragmatists, like Rinard (2015; 2017) deny it (that is why I call her approach ‘radical,’ see ch. 3.2 for how her view contrasts with the view of more moderate pragmatists). As I argue below in 4.2.3, Steglich-Petersen (2011) and Steglich-Petersen/Skipper (fc) also deny the normative force of object-given reasons.

Or, and this is the option I explore more thoroughly here, we *deny* that there are any other considerations that could potentially be considered as reasons for an attitude besides object-given considerations. The only way to make sense of a denial of (RC*) is then to deny that object-given considerations are reasons: they are mere considerations concerning truth, desirability, or rightness. When these considerations clearly indicate truth, desirability, or rightness, then this does not say anything about what we *ought* to, or even what we have *reason* to. It would not be true that we always ought to respond correctly to these considerations because they lack the normative force of reasons. A ‘correct’ response to these considerations could then at most be analyzed in conditional terms or as what *seems* to be correct to philosophers who think that object-given considerations are reasons, or in terms of the attitude’s constitutive aim without understanding this aim in any normative way. If we deny the normative force of object-given considerations, and we deny that there is anything else that could be worthy of the name ‘reason for an attitude,’ then we are left with mental nihilism: there are no reasons for attitudes.

These two options – pragmatism and nihilism – give me a nice opportunity for mentioning my overall dialectic strategy throughout parts II, III, and IV: by spelling out a challenge for the normative force of object-given reasons in part II, I show how either pragmatism or nihilism can seem as attractive positions to endorse. I then show in part III how pragmatism fails. Nihilism might then seem the way to go. The only way to avoid nihilism turns out to be making the normative force of object-given reasons intelligible – i.e., by committing to mental constitutivism. Spelling out this normative force within a constitutivist framework (i.e., without committing to pragmatism) is the challenge of making mental normativity intelligible, thereby averting the threat from mental nihilism.

There will be overall two arguments that I consider as calling into doubt the normative force of object-given considerations: the argument from blameworthiness (spelled out in this chapter and the last one) and the argument from cases of conflict (spelled out in chapter 6). Both can be read as either supporting pragmatism or nihilism. Since I deny pragmatism in part III of this book, I take them to support nihilism. Yet I view them as *challenging* the very idea of mental normativity, rather than refuting it. For we should not endorse a radical claim like nihilism before having considered ways of replying to the two arguments.

What does this focus on reasons help us when we want to solve the problem of mental responsibility? Remember that the purpose of our excursion into normativity is to evaluate the plausibility of *The easy solution* (cf. chapter 0.1). This solution was problematic because our attitudes are different from mere sensations in that they are within the space of reasons. If we

deny that they are ever held for reasons, our attitudes would be on a par with our sensations, and our responsibility for them would be as indirect as *The easy solution* claims. We saw that, maybe surprisingly, there is indeed a case to be made for the radical claim that there are no reasons for attitudes. Yet we also saw some reasons for being suspicious about this radical claim (besides the fact that it is radical): First, we sometimes seem to relate to people who are epistemically irrational in a way that presupposes that they are responsible for their irrationality, namely in the form of epistemic distrust (cf. chapter 4.3). Secondly, we saw that endorsing *The easy solution* with respect to beliefs, as consequential conceptions of doxastic responsibility do, faces the dilemma of either endorsing nihilism or denying a very minimal and plausible connection between reasons and blameworthiness (premise (1) of the argument from blameworthiness), thereby committing to anti-responsibilism, which also threatens to collapse into nihilism (cf. chapter 4.4). Finally, we had trouble arguing for the idea that there are no reasons for certain emotions: we seem to blame people in a sense for failing to respond correctly to their reasons for emotion, at least when they fail so regularly (cf. chapter 5.1.5).

I suggest that we pursue these limitations of *The easy solution* further by continuing to explore the status of object-given reasons for attitudes. Before we do so, I wish to make us aware that the concept of ‘normativity’ is a philosophical artefact with different, but related, uses. In defending the normativity of rationality, we should thus be aware of what it is we want to defend.

5.3 WHAT IS NORMATIVITY?

In chapter 0.2, I introduced the idea of normativity by giving examples of *non*-normative standards like the following:

- (a) ‘An oak should have deep, sturdy roots.’
- (b) Rules of etiquette, like ‘One ought to use fork F for meal M.’

My preliminary understanding of normativity was that these non-normative standards do not provide us with reasons, while the latter do. Yet, as I will point out below, this understanding of normativity does not allow for a distinction between two *kinds* of normative reasons. This is because in this sense of ‘normative,’ both reasons for action and reasons for attitudes are normative reasons in the same sense according to my account. In order to make a distinction *within* the category of normative reasons (in this preliminary sense), it makes sense to look at what other concepts of normativity are employed in the philosophical discussion and see

whether some of them are only applicable to reasons for action. I will propose in this subchapter that we should acknowledge a wide range of concepts of normativity, thereby allowing for ways in which rationality can be said to be normative, and ways in which it is not. My reason for this proposal is that the concept of normativity is a philosophical artefact without any corresponding use in ordinary language. It is thus up to us how we want to use it, where this use is only restricted only by considerations concerning how well this use serves our theoretical purposes.

In this book, I will defend the normative force of object-given reasons, or the standards of rationality, in some senses I distinguish below, but not in others. I will defend the normativity of these reasons in the sense that, firstly, they can be considered as full-blooded *reasons* (in a sense), secondly, that non-compliance with them can give rise to forms of blame (to which I will, in part IV, refer to with what I will call ‘the broad concept of blame’), thirdly, that they explain the truth of specific ‘ought’-statements, and, finally, that they *normally* motivate an ideally rational agent to comply with them.

I will *not* defend the normativity of these reasons, firstly, in the sense that defines normativity terms of value, secondly, the sense in which the violation of normative standards can give rise to what I will later call ‘passionate blame’ and, thirdly, the sense in which normative reasons are reasons that *always* motivate the ideally rational agent to comply with them. These latter concepts of normativity might all be applicable to reasons for action, but not to reasons for attitudes. I summarize this in 5.4.

5.3.1 NORMATIVE STANDARDS AND REASONS

The first way of explaining why (a) and (b) are non-normative while, say, the requirements of morality are normative, is to say that both (a) and (b) do not imply reasons, while the requirements of morality do. That is, even if (a) or (b) were true, it might still be the case that you do not have a reason to ensure that oaks have deep, sturdy roots, or, respectively, that you do not have a reason to use fork F for meal M. According to Joseph Raz, ‘all normative phenomena are normative in as much as, and because, they provide reasons or are partly constituted by reasons’ (2011, 85). Kieseewetter also chooses this way for distinguishing normative standards from non-normative standards when saying the he is

using the term ‘normative’ in a [...] sense, in which it refers not to just any standard or requirement, but only to those that are necessarily accompanied by *reasons* for conformity. Many requirements are not normative in this sense. For example, the rules of the German card game Skat require me to

distribute the cards in the order 3, 4, and 3, and the Chicago Manual of Style requires me to use the Oxford comma, but I may have no reason to conform to these requirements. I can accept that the standards in question exist and still sensibly ask whether I should conform to them. Drawing on Christine Korsgaard, Broome calls this question ‘the normative question.’[...] To ask the normative question is not to ask whether there is some standard, according to which some course of action counts as correct. It is to ask whether a given standard is authoritative or binding for a person. And this, I think, is best understood in terms of normative reasons (Kiesewetter 2017, 3).

Just because oaks *ought to* have deep, sturdy roots, or just because I *ought to* deal the cards in a certain order (when playing Skat), or just because I *ought to* use the Oxford comma (according to the Chicago Manual of Style), it is not necessarily the case that I have a *reason* to do so. Some ‘oughts’ imply reasons, some do not. Only the former are *normative* standards.

This characterization of normativity will not suffice for my argumentative purposes. It does not allow for a distinction between standards of rationality and requirements of morality or prudence. According to the position I advance, epistemic norms and other mental norms provide us with reasons and make it true that we (in some sense) ought to have certain beliefs and other attitudes, but the violation of epistemic and other mental norms has normative consequences that are different from the normative consequences of violations of the requirements of prudence or morality. At the end of part II (‘Summary and Conclusion of Part II’), I will introduce two distinct senses of ‘ought’ and ‘reason.’ Since I think that there are two such senses, Raz’s and Kiesewetter’s explication of the concept of normativity is not sufficient for my purposes. For according to my position, even if we grant that all normative standards provide us with reasons, they do not provide us with reasons in the same sense of ‘reasons.’ Probably there are other concepts of normativity that help me distinguish these senses of ‘reasons’ – senses of ‘normativity’ in which it turns out that some reasons are not normative reasons (in these senses of ‘normative’).

5.3.2 NORMATIVE STANDARDS AND VALUE

Another possible answer to why (a) and (b) are not normative might allow for a distinction between two concepts of ‘reason.’ According to this answer, sentences like (a) and (b) do not imply anything about how *valuable* it would be to follow the expressed norm in a concrete situation. This answer distinguishes normative standards from non-normative ones by saying that the former are placed on us because what is brought about by complying with the standard is valuable, while the latter are not placed on us because of the value of the consequences of

compliance. The reason I have in a situation not to lie to somebody might, for example, be provided by the value of the friendship I would endanger by lying to the person. The reason I have in a situation not to harm somebody might be provided by the disvalue of the pain I would cause. The requirements of morality are normative precisely because the reasons they provide are grounded in the value of the consequences of actions. By contrast, when we comply with the norms of etiquette, we do not thereby necessarily promote value.¹⁶

This concept of normativity straightforwardly leads to the claim that the standards of rationality are not normative. For complying to these standards will not always promote value (e.g., in cases in which one is better off when believing against the evidence). The reasons provided by the rationality of attitudes are not grounded in value in the way reasons for action seem to be grounded in value. However, evaluating whether reasons for action indeed differ from reasons for attitudes in that they are exclusively grounded in value would require an extensive discussion beyond the scope of this inquiry. If they were, then this would allow reasons for action to be ‘normative’ in a way in which reasons for attitudes are not, and thus back up my claim that reasons for action are reasons *in a different sense* than reasons for attitudes are ‘reasons.’¹⁷

5.3.3 NORMATIVE STANDARDS AND RESPONSIBILITY

The third possible answer to why (a) and (b) are not normative is that nobody is *responsible* to standards that say that an oak should have deep sturdy roots, or that one should use fork F for meal M.¹⁸ If an oak fails to be in line with (a), or if I fail to conform to the rules of etiquette, nobody is to be *blamed* for this. By contrast, if my reasons for not harming someone are *decisive*, I will, at least in the absence of excuses, be blameworthy if I do not comply with them. This fact shows that the considerations that count against harming the person have normative force. Not using fork F for meal M does never by itself imply that one is blameworthy for doing so. This is because one is blameworthy for not doing so only if there are prudential or moral

¹⁶ When Kolodny talks about ‘substantive reasons’ (2005, 554-5), he might have in mind reasons which are provided by value. For his examples for substantive reasons for X-ing are ‘that X-ing would prevent suffering, or advance the frontiers of knowledge’ (ibid.).

¹⁷ Raz is a famous proponent of this understanding of reasons for action in terms of value. Cf. Raz 2011, 41-5 for a discussion of value-related differences between reasons for actions and reasons for beliefs.

¹⁸ On the notion of ‘being responsible to a norm,’ cf. ch. 0.2.3.

reasons to comply with the rules of etiquette (as there might not be if, for example, one eats with people who defy etiquette).¹⁹

Since I will endorse a hybrid account of mental responsibility in part IV of this book, my account allows me to distinguish types of normative reasons. Reasons for attitudes, according to this conception, are normative *in a broader sense* because they are associated with the broader concept of blame. In this sense of ‘blame,’ also responses like distrust or judging that one’s interpersonal relationship is impaired can count as blame. To be blameworthy *in the narrow sense* – i.e., to be legitimately subject to responses like resentment and indignation –, however, presupposes that one failed to comply to requirements of voluntary conduct. Failing to respond correctly to reasons for action can thus give rise to blameworthiness in the narrow sense, while failing to respond correctly to reasons for attitudes can only give rise to blameworthiness in the broad sense. This is the way I prefer to distinguish different types of normative reasons.

Thus, depending on what we mean by ‘normativity,’ the standards of rationality will turn out to be normative or not.²⁰ Since the concept of normativity is a philosophical artefact, we should be clear about what we mean by the term, rather than argue about which concept of normativity is ‘the right one.’ According to my position, the standards of rationality are normative in some senses of the word, but not in other senses. The norms of rationality will provide us with a certain type of reasons, and non-compliance with these norms can make us sometimes, in a sense, blameworthy. In this sense, rationality is normative. Yet complying with the standards of rationality does not promote value, as the pragmatist claims. Nor does non-compliance with the standards of rationality legitimately give rise to passionate blaming responses.

5.3.4 CONCEPTS OF NORMATIVE REASONS

There are definitions of ‘normative reasons’ according to which object-given reasons for attitudes will turn out to be normative reasons, and according to which they will not.

¹⁹ For such a conception, cf. Kauppinen 2018, 3. One potential worry one might have with this characterization of normativity is that the explanatory direction should be the other way around: rather than understanding normativity through blameworthiness, we should understand blameworthiness through normative reasons. I do not see, however, why it is not possible to illuminate both notions by defining them reciprocally. After all, both ideas are very central to our ordinary understanding of our moral practice, and so it seems that we gain some understanding of each if we express their conceptual relation among another.

²⁰ See Glüer/Wikforss (2018) and Maguire (2018) for some recent doubts about the ‘normativity’ of reasons for attitudes. For Maguire, reasons for attitudes are merely ‘fit-making facts.’ According to my position, there are good reasons to call reasons for attitudes ‘reasons:’ they are conceptually connected to responsibility and blame.

Wedgwood (2017, 89-96) distinguishes between concepts of normative reasons as ‘normative-explanation reasons’ and ‘ideal-motivation reasons,’ both of which are pervasive in the literature on normativity.²¹ Take normative-explanation reasons first. These are facts that explain why certain ‘ought’-statements are true. It seems obvious that reasons for attitudes can explain the truth of certain ‘ought’-statements (about what an agent ought to believe, desire, feel, intend). There are ‘ought’-statements that are made true merely by the presence of *decisive* reasons for attitudes – that is, reasons which make a response that is supported by them the only correct response to these reasons.²² I do not deny that responding to decisive reasons for an attitude by forming that attitude is a *correct* response to these reasons, and so I do not deny that reasons for attitudes are normative in this sense. Yet there is a sense in which we *can* ask the further question of whether we *ought* to give a correct response to our reasons for attitudes. Thus, I commit myself to the existence of a further concept of normativity, or a use of ‘ought,’ in which we can sensibly ask this question. I will later refer to this ‘ought’ as the *voluntary ought* (cf. ‘Summary and Conclusion of Part II’).

Consider, next, the concept of normative reasons as ‘ideal-motivation reasons.’ These are facts that would motivate a fully informed and rational agent to respond to them. Again, one might think that if reasons for attitudes are *decisive*, then they would motivate a fully informed and rational agent to respond to them by forming the attitude supported by those reasons. However, this is not always true. Suppose that, in a specific case, you ought to *cause* yourself not to respond correctly to your decisive reasons for an attitude – for example, a case in which you ought to cause ourselves not to properly base your beliefs on your evidence. In this case, an ideally rational agent would not respond correctly to his epistemic reasons, but rather cause themselves *not* to do so. Thus, reasons for attitudes are not normative in the sense that they would always motivate an ideally rational and informed agent to respond correctly to them.²³ We could at most say that, *normally*, ideally rational agents are disposed to conform to the standards of attitudinal rationality.

²¹ As a proponent of the first understanding, Wedgwood cites Broome (2004). As proponents of the second, he cites Setiya (2007, 12), Kearns and Star (2009), as well as Williams (1995, 35).

²² Note that the notion of ‘correctness’ I use is different from Wedgwood. For Wedgwood, an attitude is correct if it fulfills its constitutive standard (truth, desirability, rightness, etc.). When I talk about a ‘correct’ response to a set of reasons I mean the response that these reasons favour (a use common in the current debate about rationality, cf. ch. 0.2.4). For how nihilists and pragmatists might use ‘correct response,’ cf. 4.2 above.

²³ That there are cases in which an ideally rational agent would not respond in the way it is correct to respond supports Wedgwood’s claim that the two concepts of normative reasons he introduces rely on different notions of ‘reason’ (2017, chs. 4.4-4.6).

5.4 SUMMARY

I started out in this chapter by extending the argument from *doxastic* blameworthiness to an argument from *mental* blameworthiness, thereby setting up a challenge for the idea that there are any reasons for attitudes. In the course of my discussion, we saw that the premise that we cannot be blameworthy merely for non-compliance with object-given reasons did not seem plausible when we considered the emotions of admiration and love. This should motivate us to look closer at the kind of blameworthiness at play here, and to see whether this kind of blameworthiness can arise from purely epistemic failures and other defects of rationality besides emotional irrationality.

I think blameworthiness can arise from irrationality, and I will return to this at the end of chapter 6 as well as in part IV. Evaluating *The easy solution* and thereby responding to the problem of mental responsibility will not require me to spell out the various forms of blameworthiness and criticizability that the violation of the standards of rationality can give rise to. Yet it will require me to sketch a plausible general account of this kind of blameworthiness, thereby highlighting the distinction between rational standards and practical requirements of prudence or morality that provides my framework for solving the problem of mental responsibility: we are responsible to these two types of norms in different ways.

Subchapters 5.2 and 5.3 were mainly clarificatory. We saw that if we want to preserve the normativity of rationality, we have to specify in what sense we want to do so. I am interested in rationality insofar as we understand being rational as responding correctly to our object-given reasons for attitudes. For I am interested in the *normative force* of object-given reasons for attitudes. Saving their status as a reason, I have argued throughout the last two chapters (especially in chapters 4.1, 4.2, and 5.1.1), requires us to make sense of how we can be blameworthy or criticizable merely in virtue of the fact that we fail to respond correctly to our object-given reasons. Since anti-responsibilists (cf. chapter 2.6) deny such a connection between blameworthiness and reasons, their position fails.²⁴

Making sense of the idea that rationality provides us with reasons thus requires us to make sense of the idea that we can be blameworthy for non-compliance with the standards of rationality. The first and the third concept of normative standards I distinguished (chapters 5.3.1 and 5.3.3) thus go hand in hand as long as we understand blame as what I will later call ‘the broad concept of blame’, i.e., blame as the marker of impaired relationships rather than

²⁴ My argument in chapter 11 that we are, in a sense, directly responsible for attitudes, will provide a further argument against anti-responsibilism.

understood narrowly as merely involving passionate emotions like resentment or indignation. Furthermore, these are standards that provide us at least with ‘normative-explanation reasons.’ They also provide us with ‘ideal-motivation reasons’ if the latter are understood as *usually* motivating an ideally rational agent (i.e., ignoring cases in which the agent ought to cause their own non-compliance). This is the sense in which I think the norms or standards of rationality are normative. We will see why they nevertheless do not provide us with reasons in the same sense of ‘reasons’ in which requirements of morality provide us with reasons if we consider another argument that is challenging the normative force of object-given reasons, and thus the normativity of rationality: the argument from cases of conflict.

Chapter 6

The Argument from Cases of Conflict

The present chapter introduces another argument to the conclusion that there are no reasons for attitudes (6.1-6.3). The purpose of discussing this argument is twofold. First, it brings out the challenge of understanding mental normativity, arising from the threat of mental nihilism, more clearly. I will have to partly meet this challenge in order to evaluate the plausibility of *The easy solution* and to thereby reply to the problem of mental responsibility.¹ Furthermore, discussing this new argument will allow me to initially present the outlines of my reply to our challenge of understanding the normative force of object-given reasons, and thereby also the outlines of my reply to our main challenge of solving the problem of mental responsibility in response to the argument that I discuss here (6.4). I summarize the dialectic after this chapter in a separate section that concludes part II of this book, where I also return to the argument from blameworthiness that I discussed throughout chapters 4 and 5.

6.1 DEONTIC TRANSMISSION BETWEEN RESPONSES

Consider the following principle:

Deontic action-belief transmission (ABT).

If you ought to cause yourself not to believe *p*, then it is false that you ought to believe *p*.

If you ought to cause yourself to believe *p*, then it is false that you ought [not to believe *p*].

If (ABT) was true, this would have interesting – and yet puzzling – consequences for the normativity of epistemic reasons. For if the cases mentioned in the antecedent condition of (ABT) are possible, then we have a *prima facie* plausible argument to the conclusion that we ought not always be epistemically rational. To illustrate, assume that ‘*p*’ is an epistemically rational belief:

(1) *S* ought to cause themselves not to believe *p*. (*cases of conflict are possible*)

¹ As a reminder: I only have to respond partly to this challenge insofar as I do not have to provide detailed analyses of the various kinds of mental normativity – i.e., I do not have to explain in detail which kinds of reactions are appropriate towards someone’s irrational belief, desire, emotions, intentions, etc. Rather, I just need the plausible sketch of a general account of such reactions which I will provide especially in part IV, but which I will already discuss in ch. 6.4 below.

(2) If S ought to cause themselves not to believe p, then it is false that S ought to believe p.

(ABT)

(3) So, it is false that S ought to believe that p.

Since ‘p’ refers to an epistemically rational belief, (3) states that it is false that S ought to be epistemically rational. That cases of conflict are possible, as (1) states, should not be controversial. The possibility of cases of conflict arises from the gap between theoretical and practical rationality. To be theoretically rational is to properly base one’s beliefs on one’s evidence, i.e., to respond correctly to one’s (object-given) reasons for belief.² To be practically rational is to intend or to do what is right. Since what action is right is at least partly a matter of how valuable the available actions and its consequences are, and since responding correctly to one’s reasons for belief is not always valuable, we can conceive of cases in which we should cause ourselves not to respond correctly to our reasons for belief. Pragmatists usually point to such cases in defending their position: cases where believing what is epistemically rational is not the best available option, and where we thus should *cause* ourselves to believe – or, according to the pragmatists, just *believe* – against our evidence. Thus, I will not further discuss (1) in what follows.³

What is crucial is premise (2), i.e., (ABT). We can generalize (ABT) and the argument from above easily to *all* mental states, thereby not merely bringing out a conflict between theoretical and practical rationality, but rather a conflict between mental normativity and the normativity of action more generally:

² It is important to remember that my use of ‘to be rational’ is always short-hand for ‘to respond correctly to one’s (object-given) reasons for an attitude.’ I did not defend a reasons-based conception of rationality in this book, like Kiesewetter (2017) or Lord (2018) did. Yet the phenomenon I am interested in is the same as the one they argue rationality consists in. Since the challenge I continue spelling out here questions the very existence of object-given reasons for attitudes (cf. chs. 4 and 5), the ‘correctly’ in ‘responding correctly to one’s reasons’ might seem dubious. Yet it can be provided with an epistemic or conditional reading in order not to undermine the nihilists position from the very outset (cf. ch. 5.1.1). For if it turns out that object-given reasons are, after all, no reasons, then there is no sense in which we *ought* to respond to them: We *ought* to respond only to reasons.

³ Hazlett (2013) presented a detailed treatment of the value of true belief and argued, among other things, that it is sometimes better to have beliefs that are false, or not epistemically justified (cf. Hazlett 2013, e.g., 44-52). This allows for cases in which we should bring ourselves to have false or epistemically unjustified (irrational) beliefs – i.e., ones that are either incoherent with our other beliefs or which are not sufficiently supported by reasons in order to count as *correct* responses to your reasons for belief. The ‘causing’ of irrational belief might involve practices of ignoring evidence, self-deception, or even cultivating systematic unresponsiveness to certain evidence.

Deontic action-attitude transmission (AAT).

If you ought to cause yourself not to have attitude M, then it is false that you ought to have M.
If you ought to cause yourself to have attitude M, then it is false that you ought [not to have M].

‘To have M’ can here be read shorthand for ‘to believe, desire, feel, intend, ...’. (AAT) states that the deontic status of a certain *attitudinal* response sometimes depends on the deontic status of an *action* of causing oneself to give this response. If you are required to cause yourself not to give an attitudinal response, then it is false that you are required to give the response. This implies that a required action that results in an attitudinal response *transmits a permission* to the attitudinal response in which the action results.⁴

The truth of (AAT) would have implications for current philosophical debates. Consider, first, the question of whether *evidentialism* is true: is it true that we always *ought* to respond correctly to our epistemic reasons?⁵ It seems that we can imagine scenarios in which we ought to *cause* ourselves not to be epistemically rational (cases in which believing against our evidence might improve our chances of survival, or our chances of success, or cases in which an eccentric billionaire offers us a lot of money for believing against our evidence). If (AAT) was true, then it would not be true that we ought to be epistemically rational in these cases, and evidentialism would turn out to be false.⁶ Thus, evidentialists are committed to denying (AAT).⁷

Second, (AAT) would be devastating for the idea that rationality is *strictly* normative – that we always ought to be rational. In the recent philosophical literature, rationality is either understood as mental coherence, or as responding correctly to one’s reasons for attitudes (see

⁴ To see how (AAT) implies a transmission of a permission, consider the following implication: If it is false that you ought to ϕ , then it is allowed for you not to ϕ . Thus, (AAT) implies: If you ought to cause yourself not to ϕ , then it is allowed for you not to ϕ .

⁵ ‘To respond correctly to one’s epistemic reasons’ means to properly base one’s beliefs on one’s evidence. For a discussion of what it could mean to *properly* base one’s beliefs on one’s evidence if one is a nihilist, cf. ch. 5.1.1.

⁶ Kiesewetter (2017, 180-5) argues that epistemic irrationality implies that one fails to properly base one’s beliefs on the evidence to which one *attends*. This won’t affect the argument against evidentialism I present here. For we might well conceive of cases where we have the means to believe against the evidence to which we attend – though these scenarios will tend to be more unrealistic and involve, for example, futuristic devices by means of which we can manipulate our beliefs. Insofar as we tend to respond rationally, it will be difficult not to respond to evidence to which we attend.

⁷ Rinard (2015, 123) uses a similar principle she calls ‘equal treatment’ for arguing against evidentialism, and in favour of her version of pragmatism. I say that the principle is ‘similar’ because I think equal treatment implies (AAT): if the normativity of attitudes is to be understood like the normativity of any other state (which is the claim of equal treatment), then there would be no grounds left for denying (AAT). As I argue below, the principle (AST) which is concerned with ordinary states rather than with reasons-responsive attitudes, is very plausible.

chapters 0.2.4; 5.2). It seems both conceivable that we ought to *cause* ourselves not to have coherent mental states, as well as that we ought to *cause* ourselves not to respond correctly to our reasons for attitudes. If (AAT) was true, then the possibility of these cases implies that (attitudinal) rationality cannot be *strictly* normative – it is false that we always ought to be rational.⁸

Third, I will argue (in chapter 6.3) that it follows from (AAT) that (attitudinal) rationality is not even *pro tanto* normative (in the relevant sense): in a sense of ‘reason,’ we do not even have a *pro tanto* reason to be rational.⁹

These conclusions are unclear as they stand and should not unsettle us too much. Rather, we should understand these conclusions as challenging us to specify the sense in which (AAT) is true, and thus in which sense reasons for attitudes have and do not have normative force, and in which sense we can and cannot be said to be directly responsible for our attitudes. This is the challenge we accept when we do not want to commit to nihilism.

I will argue that my argument is best understood as establishing an asymmetry between reasons for attitudes and reasons for action with respect to their normative force, and thus call into doubt an equal treatment of these reasons. The idea that attitudes can be normatively required in a similar way as actions can be normatively required is a long-standing dogma within contemporary philosophy that should be rejected. There are no *requirements* of attitudinal rationality. Rather, attitudinal rationality sets up standards the violation of which has specific normative consequences concerning our interpersonal relationships – consequences that we can legitimately mark with criticism or blame, and which are appropriate even if we were in a case of conflict, i.e., even if we were *required* to cause ourselves to be irrational. While acting wrongly can sometimes have similar normative consequences, there are normative consequences in our interpersonal relationships that presuppose wrong action, and which can legitimately be marked by forms of blame that are not made appropriate merely by *being* in certain state – even if the state is irrational. These specific forms of blame are *passionate* blaming-responses like resentment, indignation, or guilt. A fuller account of this picture I sketch and the end of the present chapter will have to wait until part IV.

⁸ One might doubt that we can conceive of scenarios in which we ought to cause ourselves not to respond correctly to one’s reasons for *intention*. I will argue that we can do so in ch. 6.3 below.

⁹ For the relevant senses of normativity to which I return in this chapter, cf. ch. 5.3. Barry Maguire (2018) recently argued, on other grounds than I present here, that reasons for affective attitudes – like desires, emotions, or hope – are no normative reasons, but rather what he calls ‘fit-making facts.’ With respect to reasons for belief, a similar view was articulated by David Papineau (2013). I suspect that Susanna Rinard, who seems to endorse (AAT) (see fn. 7 above), would agree that any object-given reasons for attitudes are not normative.

I start by providing an argument in favour of (AAT) (6.1). I discuss what I take to be the most intuitive objection against this argument: the two-domains reply. The defender of (AAT) can, I argue, present a powerful response to this reply that makes (AAT) seem highly plausible. I then argue that we can construct a convincing argument to the conclusion that if it is false that we always ought to be rational (i.e., if it is false that we always ought to respond correctly to our object-given reasons for attitudes), then it is also false that we always have a *reason* to be rational: the *strict* normativity of rationality stands and falls together with the *pro tanto* normativity of rationality (6.2). This puts the mental nihilist in a good position.

I then turn to an objection from the criticizability of irrationality and thereby clarify the import of my discussion (6.3). Rather than rejecting the normativity of rationality altogether on the grounds that there are no reasons for attitudes, I argue that we should acknowledge that the normative force of reasons for attitudes is to be understood in different terms than the normative force of reasons for action. I point out how this conclusion might help us solving the puzzle of mental responsibility. Part IV of this book will provide further support and clarification to the distinction between requirements of action and norms of attitudes I make here.

6.2 AN ARGUMENT AGAINST THE STRICT NORMATIVITY OF OBJECT-GIVEN REASONS

(AAT) and the possibility of cases of conflict together allow for an argument against the strict normativity of object-given reasons: If (AAT) is true, and cases of conflict are possible, then it is not true that we always *ought* to respond correctly to (what we would normally regard as) decisive object-given reasons for an attitude. I discuss (AAT) in detail in the present subchapter to evaluate this argument, before turning to its implications for the *pro tanto* normativity of object-given reasons and then to my reply to the challenge posed by the argument.

6.2.1 FROM ORDINARY STATES TO REASONS-RESPONSIVE ATTITUDES

Consider yet another principle that is similar to (AAT), but much less controversial. Take ‘s’ to refer to any ordinary state of the world (or of a person) that is not directly responsive to reasons, like the state of ‘being outside’:

Deontic action-state transmission (AST):

If you ought to cause yourself not to be in state s, then it is false that you ought to be in s.

If you ought to cause yourself to be in s, then it is false that you ought [not to be in s].

To illustrate the plausibility of this principle, imagine someone tells you in the morning ‘You ought to go outside today to see our garden.’ Later the person asks you if you have been outside. After your affirmative reply, this person accuses you: ‘Why have you been outside?! I merely allowed you to *go* outside, not to *be* outside!’ This would surely be puzzling, for going outside (used as success term) *implies* being outside. In this case, the person required you to cause yourself to be outside by telling you to go outside. In telling this, the person allowed you *be* outside (at least for a while): it would be inconsistent for them to think both that you ought to *be* inside all day, and that you ought to *go* outside at one point. Thus, (AST) seems very plausible.¹⁰

On the assumption that reasons-responsive attitudes, like believing, desiring, feeling, or intending, are to be treated analogously to non-reasons-responsive states, (AST) implies (AAT).¹¹ To be sure, this assumption is controversial. Yet we need to spell out *why* it is controversial. The most obvious objection is to point out that attitudes are states *that are responsive to reasons*, and that they thus exhibit a specific kind of normativity that renders (AAT) implausible. According to this objection, even though (AST) is not disputable, (AAT) is implausible because it is concerned with reasons-responsive attitudes rather than ordinary states. I will now turn to spelling out this objection from the two domains, and then present a powerful reply of the defender of (AAT). Understanding and evaluating this reply properly is, I argue, the key to understanding the difference between attitudinal norms and requirements to perform specific actions.

6.2.2 THE TWO DOMAINS REPLY

Evidentialists are prone to reply against (ABT) that beliefs are essentially subject to epistemic rationality. What distinguishes ordinary states from attitudes is that the latter, but not the former, are responsive to reasons. This difference allows us to accept (AST) but deny (ABT)/(AAT).

¹⁰ Note that possibilism is not a problem for (AST). Possibilists claim that even if we know that we will perform a certain wrong action B when choosing to do A, it can still be that we are required to do first A and then non-B insofar as we still *could* do non-B (cf. Jackson/Pargetter 1986, who argue against possibilism). Thus, one might think even if we *will* be irrational when we cause ourselves to be so, it can still be true that we are required to be rational in this case insofar as we still *could* be rational. However, the cases are not analogous. For the term ‘to cause’ is used as a success term. It is not possible to cause oneself to be in a state without then being in that state as a result of one’s causing. Thus, there is no sense in which I still *can* be rational if I caused myself not to be. Thus, possibilism does not pose a threat to (AST). Thanks to Konstantin Weber for pointing out this objection to me.

¹¹ Cf. Rinard’s (2017) argument of ‘equal treatment’ for her version of doxastic pragmatism.

According to this reply, whether we ought to believe something is wholly independent of whether the belief would be practically valuable or of whether we are practically required to have the belief (cf. Kelly 2003, 619; Wedgwood 2017, 41-52). We can see this independence if we specify the meanings of ‘ought’ in cases of conflict:

(C) *S* *practically* ought to cause themselves to be a way they *epistemically* ought not to be.

If we are confronted with this correct interpretation (C) of cases of conflict, then we see – the evidentialist argues – that (ABT) is obviously false. For then it reads:

(ABT*) If you ought to *practically* cause yourself not to believe that *p*, then it is false that you *epistemically* ought to believe that *p*.

Since, according to the evidentialist, the epistemic domain is the only relevant domain for evaluating what beliefs we ought to have, (ABT*) is the only meaningful reading of (ABT). Since (ABT*) is obviously false, (ABT) is false as well.

This reply can be generalized to all attitudes. We begin by pointing out that all attitudes have their constitutive aims: we ought to desire what is desirable, fear what is dangerous, admire what is admirable, intend what is right, etc. We then argue that to say that we ought to have a specific attitude can only mean to say that we ought to have attitudes that fulfill their constitutive aim, or that we ought to be rational and thereby ‘do’ our best in order to ensure that we reach an attitude’s constitutive aim.¹² This step is plausible if we do not assume pragmatism about an attitude’s rationality.¹³ But attitudes have constitutive aims or are subject to rational norms quite independently of whether we ought to *cause* ourselves not to reach them or comply with them. Thus, the only meaningful reading of (AAT) is:

(AAT*) If you *practically* ought to cause yourself not to have an attitude *M*, then it is false that you *constitutively/rationally* ought to have *M*.

However, if we acknowledge the distinctness of constitutive aims or rational norms from requirements of action, we see that (AAT*) is clearly false. Of course we can sometimes be required to cause ourselves to have an attitude that is not fully rational or does not reach its

¹² It might be rational to believe that *p* and yet believing *p* might not fulfill belief’s constitutive standard of truth. Whether one thinks that we always ought to reach an attitudes constitutive aim or whether we always ought to be rational (and thereby make it most likely that we reach that aim) will be a matter of whether one accepts objectivism or subjectivism about the mental ‘ought.’ I do not decide this question in this book. On objectivism and subjectivism, cf. ch. 4.2.2.

¹³ On pragmatism, cf. ch. 3 as well as my extensive discussion throughout part III.

constitutive aim, as the proponent of (AAT) has to accept to get their argument against the normativity of rationality running.

6.2.3 RESPONDING TO THE TWO-DOMAINS REPLY

Yet the two-domains reply will not convince the proponent of (AAT). The proponent will argue that there must be another meaningful reading of (AAT) than (AAT*), and they will call into question whether there can be any kind of requirement or normative standard in place that would conflict with the practical requirement in cases of conflict. Consider the following, powerful, line of argument against the existence of an epistemic standard in cases of conflict as they are relevant to (ABT):

- (1) In cases of conflict, you cannot [do the right thing *and* believe the right thing.] [*conceptual truth*]¹⁴
- (2) ‘Ought’ implies ‘Can.’ [*plausible principle*]
- (3) Thus (from (1), and (2)): in cases of conflict, you it is false that you ought to [do the right thing *and* believe the right thing.]
- (4) In cases of conflict, you ought to cause yourself not to be epistemically rational. I.e., you are *required to do the right thing*. [*conceptual truth*]
- (5) If it is false that you ought to [ϕ and ψ], but you ought to ϕ , then it is false that you ought to ψ . [*plausible principle*]¹⁵
- (6) Thus (from (3), (4), and (5)): in cases of conflict, it is false that you ought to believe the right thing, i.e., it is false that you ought to believe what is epistemically rational.

The argument is powerful because each premise seems to be highly plausible, and it is thus hard to pin down where and why the argument goes wrong. Let me comment on the premises in turn to highlight their plausibility. I will focus on the argument as stated – as supporting (ABT) rather than the more general principle (AAT). It is easy to see that the argument can be generalized to all mental states, and so I will lead the following discussion without formulating this generalized version of the argument.

Premise (1) is just an implication of the fact that cases of conflict are possible. In cases of conflict, we *will* flout at least one norm when we comply with the other. When we do the

¹⁴ ‘To do the right thing’ is just an abbreviation for ‘to cause yourself not to be epistemically rational.’ ‘To believe the right thing’ is just an abbreviation for ‘to believe what is epistemically rational.’

¹⁵ Kiesewetter discusses a similar principle under the label *Agglomeration* (2015, 930-6; 2016, 6).

right thing, we cause ourselves not to believe the right thing. When we believe the right thing, then this means that we did not do the right thing – i.e., did not cause ourselves not to believe the right thing (for ‘to cause’ is used as implying success). We cannot both respond correctly to our reasons for actions and to our reasons for belief. That is just what cases of conflict are about.

Premise (4) is another implication of the possibility of cases of conflict. For cases of conflict are *not* to be conceived as problematic dilemmas. When confronted with such a case, we are reacting to the overall situation in the right way by *doing* what we ought to (that is, if we are causing ourselves not to be epistemically rational) rather than by being epistemically rational and thereby violating the practical requirement. There is no question that the practical requirement is in place: practical requirements do not get disabled just because one of their consequences is that an epistemic standard is violated.¹⁶

In what sense, then, should it also be true that we (normatively) *ought* to be epistemically rational in a case of conflict, if, at the same time, we ought to *cause* ourselves not to be epistemically rational? To reach the conclusion that there is no such sense in which we also ought to be epistemically rational, we need to assume ‘Ought implies Can’ (premise (3)) and another plausible principle (premise (5)).

Note that the ‘Ought implies Can’ that we need here is not controversial at all. For this principle merely has to state that it must be *conceptually* possible to ϕ in order for it to be possible that ϕ ing is what we ought to. This is because in cases of conflict, it is not merely physically impossible to [do the right thing and believe the right thing], but rather *conceptually* impossible.

The crucial premise might indeed be (5). Yet (5) is highly plausible when we think about actions. Suppose you are *not* required to [wash your clothes and go to the laundry]. Now assume that you are required to wash your clothes. Since, according to the first assumption, you would be allowed to wash your clothes without going to the laundry (say, at home), it will follow, almost trivially, that you are not required to go to the laundry. For otherwise, how could you still not be required to [wash your clothes and go to the laundry]?

Yet even if (5) is plausible when we think about actions, it need not be plausible when the two responses in questions are an action and an attitude. However, if we accept (5) for actions but not for attitudes, this might call into doubt that the normative force of reasons for

¹⁶ Of course, the same might be said for epistemic norms (thereby denying the conclusion of the argument). But then one would have to reject one of the other premises of the argument.

attitudes is to be understood in similar terms as the normative force of reasons for action. I will argue in section 6.3 that this is exactly the conclusion to draw from the discussion here, and that this is why (5) is false. For now, let us see what the consequences for the normativity of rationality would be if the argument of the proponent of (AAT) was sound.

6.3 AN ARGUMENT AGAINST THE PRO TANTO NORMATIVITY OF OBJECT-GIVEN REASONS

It follows from (AAT), and the possibility of cases of conflict, that we should not always respond correctly to (object-given) reasons for attitudes. This does not refute the *pro tanto* normativity of rationality: it does not refute the idea that object-given reasons are reasons at all. However, I now argue that by claiming that we always have *a reason* to respond correctly to our object-given considerations,¹⁷ although not always *decisive* reasons, we thereby commit to either of two implausible implications: either that we can weigh reasons for attitudes against reasons for actions; or we implausibly end up with two incommensurable normative directions. We should thus, according to the argument, reject the idea that object-given considerations are reasons.

Note first that we can understand the claim that we always have *a reason* to respond correctly to our object-given considerations – i.e., to be rational – in two ways. Yet we would only preserve reasons for attitudes according to the second understanding, which also commits us to the problematic implications:

- (a) If an attitude is the only correct response to our object-given considerations, then we have a reason *to bring the attitude about*.
- (b) If an attitude is the only correct response to our object-given considerations, then we have a normative reason *to give the attitude-response*.

Claim (a) would not commit us to weigh reasons for attitudes against reasons for actions, or to incommensurable normative directions. This is because the reasons rationality provides us with according to (a) would be reasons for *actions* to bring the attitude about. These reasons could be weighed with other reasons for actions as any reasons for actions can be weighed with one another. However, (a) would at the same time not preserve the idea that there are reasons for

¹⁷ In what follows, I will talk about ‘object-given considerations’ instead of ‘object-given reasons’ because it is precisely the status of these considerations as reasons that is called into question. For how mental nihilists might spell out the notion of a correct response to object-given considerations, cf. ch. 5.1.1.

attitudes, because rationality would only provide us with reasons for action. The fact that an attitude is rational would just *imply* that there are reasons for action.

For illustration of (a), consider the attitude of *love*. According to Susan Wolf (2010), objects which are worthy of our love provide us with ‘reasons of love,’ which are not reducible either to reasons provided by self-interest or by morality. For example, if a person is worthy of your love, you have reasons to perform actions that help the person flourish. I would also argue that you necessarily have a reason *to cultivate your love* towards this person. What better reason can there be to cultivate your love for someone than that this person is worthy of your love? This fact about worthiness does not merely make your attitude of love a correct response to your object-given considerations for this attitude, but it also provides you with a reason *to bring it about* that you (continue to) love them (which can be outweighed by other reasons for action). Similarly, one might argue that sufficient evidence for a belief can provide you with a reason to bring it about that you have the belief supported by it. Since such a reason is a *pro tanto* reason for *an action*, it is different from the reasons you have for *the attitude*. This is why claim (a) does not establish that rationality provides us with reasons for attitudes.

Claim (b) would indeed preserve reasons for attitudes, but accepting (b) while denying the strict normativity of rationality would result in an implausible account of the normative situation in cases of conflict. Suppose that the fact that you have decisive evidence against a belief provides you with a reason not to believe, but that having the belief would lead to very good consequences. Assume that you can cause the belief by pressing a button. In this case, you ought to press the button and thereby cause yourself to have an epistemically irrational belief. Nevertheless, the *pro tanto* reason against believing (provided by the evidence) would still stand. That is, your evidence would still *favour* not-believing, even though it does not make not-believing the response you *ought* to give. But this is, at best, puzzling. What does this epistemic reason to believe tell you about the normative status of your belief? Why is this reason not sufficient to make it the case that you epistemically *ought* to believe?

Saying that your evidence still provides you with a reason to believe though not with a decisive one amounts to saying that you *epistemically have a reason to believe* although you *epistemically ought not to believe*. It is obscure what the normative import of this epistemic reason is supposed to be and how it should play into our deliberations in a case of conflict. It does not seem in any way relevant for us to consider this reason. The nihilist will argue that there is no normative ambiguity in a situation of conflict, no feeling that we make a compromise when we cause ourselves to believe against the evidence. Rather, by causing ourselves to be irrational we do exactly what we ought to do, and without any feelings of regret, when we end

up being epistemically irrational.¹⁸ This is indicator enough that we do not flout any reason not to believe in such a situation.

Thus, if one accepts that we should not always respond correctly to our object-given considerations for attitudes, because one accepts (AAT), and one still wants to maintain that rationality provides us always with *a reason* for the attitude (in the sense of (b) above), one commits oneself to an implausible analysis of the normative situation in cases of conflict. It thus seems that, if we accept (AAT), we should not only conclude that we *ought* not always respond correctly to our object-given considerations for attitudes, but that we have *no reason* to respond correctly to our object-given considerations for attitudes. On pain of contradiction, we have to deny that object-given considerations are themselves reasons for an attitude. How should we reply to this nihilist's conclusion?¹⁹

6.4 IRRATIONALITY AND CRITICISM

We should reply to the nihilist by thinking about the nature of the blame or criticism we deem to be appropriate towards the irrational. It is often claimed that one is criticizable for being irrational. In order to get an objection against (AAT), we would furthermore have to argue that being criticizable implies that one failed to respond to one's reasons the way one ought to have. Understanding this form of criticism is the key to understanding the normativity of rationality. It will help us see in what sense it can be possible to cause ourselves to be a way we ought not to be. Yet we should grant that we can be excused from flouting a standard of rationality when we ought to cause ourselves not to flout it – at least as long as our irrationality does not have any serious normative consequences for our interpersonal relationships. The analysis of cases of conflict that I propose thus has two elements: (AAT) is false because we can sometimes cause ourselves to be a way we ought not to be; yet (AAT) gains its plausibility because in such situations we are often excused from flouting the standards of rationality.

¹⁸ Compare the situation with an ordinary practical compromise. If we should meet an old friend on pain of missing out our favorite tv-show, we can come to a conclusion what we ought to do, all things considered. The considerations in the case considered here would be incommensurable, however. And we might understandably feel a kind of regret for, say, missing out on our favorite tv-show that evening. There is nothing to regret in the case I described above.

¹⁹ Again, another lesson from this discussion might be a radical version of doxastic pragmatism, rather than nihilism: we might conclude that reasons for attitudes are not object-given, but rather practical reasons. Remember that I assume that pragmatism is false throughout part II. I provide my argument against pragmatism in part III.

6.4.1 CRITICISM, 'OUGHT,' AND THE VIOLATION OF REQUIREMENTS

To see how a connection between irrationality and criticizability can pose a problem for (AAT), consider again a case of conflict in which you ought to cause yourself to be irrational, i.e., cause yourself to fail to respond correctly to your reasons for an attitude. According to the premises of our reply to the nihilist, we can construct the following argument against (AAT):

- (1) In cases of conflict, you should cause yourself to be irrational.
- (2) If you are irrational, then you are, at least sometimes, (personally) criticizable in virtue of being irrational.
- (3) If you are (personally) criticizable in virtue of being irrational, then you failed to respond to your reasons the way you ought to have responded to them.
- (4) Thus, in cases of conflict, you should, at least sometimes, cause yourself to fail to respond to your reasons the way you ought to have responded to them.

If it was true that being irrational implies criticizability, and being criticizable implies that one fails to respond to one's reasons the way one ought to have, then cases of conflict turn out to be cases where one ought to cause oneself not to respond to one's reasons the way one ought to respond to them. This contradicts (AAT).

In what follows, I will discuss the whether we are criticizable for being irrational in the sense required for this argument – i.e., I will discuss whether (2) is plausible. I will read (3) as employing the concept of a specific *kind* of criticism that implies that one fails to respond to one's reasons the way one ought to, thus reading (3) as an uncontroversial claim. Yet before we can evaluate (2), we need to properly understand the concept of criticism with which (3) is uncontroversial. It is important to remember that I employ 'to be irrational' as 'to fail to respond correctly to one's object-given reasons.'²⁰ In what sense can failing to respond correctly to one's object-given reasons make one criticizable or blameworthy?

Kiesewetter supports the idea that irrationality is criticizable by mentioning a case described by Nadeem Hussain.²¹ In this case, religious people openly admit that they lack sufficient evidence for their belief in god, but deny that they are irrational, because they do not think that they are criticizable. Calling them 'irrational,' both Hussain and Kiesewetter argue, would amount to criticizing them. It would be incoherent, or even inconsistent, to say to them that 'I

²⁰ Cf. ch. 0.2.4 for this terminological decision.

²¹ Kiesewetter (2017, 18-9, 117-8), Hussain (unpublished, §3). Kiesewetter here does not assume his reasons-based conception of rationality. Rather, he wants to support it. Yet we might just read 'irrational' here as 'failing to respond correctly to decisive reasons' for our purposes.

think you are irrational, but this is not meant as a criticism. I don't think there is anything wrong with being irrational' (Kiesewetter 2017, 39). Kiesewetter thus concludes that a denial of the criticizability of irrationality would amount 'to a quite radical error theory about ordinary irrationality ascriptions' (41). That is, we would have to revise our pre-theoretical beliefs about what irrationality is when we deny that irrationality is always criticizable.²²

To get the argument against (AAT) running, 'being criticizable' would have to imply that the religious people failed to respond to their reasons in the way they *ought* to have. In what follows, I will spell out this sense of 'ought' by distinguishing the kind of criticism to which the violation of this ought can legitimately give rise to from other forms of criticism. I will first introduce a form of criticism that obviously does *not* imply the violation of a normative standard. I then argue that the criticism involved in irrationality is both different from this criticism and yet does not have the same normative consequences as the violation of a requirement of morality. Still, I argue, being irrational gives rise to various normative consequences which presuppose a kind of criticizability, and there is thus a sense in which we sometimes 'ought not' to be irrational even in cases of conflict – i.e., when we ought to cause ourselves to be irrational. This motivates my claim that we should understand reasons for attitudes and reasons for action as involving two different kinds of normative relations.

6.4.2 CRITICISM, THICK CONCEPTS, AND MORAL BLAME

There are forms of personal criticism that do not imply that one violates a normative requirement or a normative standard. We criticize someone by calling them 'clumsy' or 'awkward.' It is also incoherent to tell someone that 'I think you are clumsy/awkward, but this is not meant as a criticism. I don't think there is anything wrong with being clumsy/awkward.' These 'thick concepts' (Williams 1985, 129) seem to ascribe both a descriptive and evaluative property to

²² Given Kiesewetter's theory of rationality, we get a potentially odd result in this example. For according to him, we fail to be fully rational when we do not believe what our available evidence sufficiently supports, at least as long as we attend to the evidence (on the attending-condition, see ch. 4.1.1). Clearly, the religious people are aware of the insufficiency of their evidence. It thus seems that Kiesewetter is committed to their criticizability for their religious belief, and, moreover, to claim that what they *ought* to believe is that their god does not exist. This seems like an unwelcome implication. For what if their belief is meaning-making or enhances their life in other possible ways? I am ultimately committed to the same implication. I will argue below that we need to understand the 'ought' and the kind of criticizability involved properly so as not to say something implausible (or even offensive), but that the religious people might indeed be criticizable in the relevant sense in virtue of their epistemic irrationality. On the permissibility of meaning-making beliefs which are insufficiently supported by the evidence, see McCormick (2015, 61-5). I agree with McCormick that meaning-making beliefs are all-things-considered allowed in the sense that we are sometimes permitted to bring them about even in the absence of sufficient evidence (cf. chs. 1.1, 1.2, 7.1).

the object to which they are applied. According to an influential account of thick concepts,²³ their correct application seems to presuppose that certain descriptive conditions plus *evaluative* conditions are fulfilled. ‘Being clumsy’ would mean, roughly, to lack skill in a specific way *and* to be criticizable for it.

If the criticism implied by irrationality-ascriptions was the same as the one that is implied by these thick concepts, then it would not be the right kind of criticism to get the argument against (AAT) running. For according to the relevant premise, being criticizable implies *that one fails to respond to one’s reasons the way one ought to have*. But does one have any reason not to be clumsy or awkward? ‘Having reason’ does not seem to be applicable to non-reasons-responsive states. Being clumsy or awkward are not states that are responsive to reasons at all – in contrast to states of believing or intending. Having reasons for not being clumsy or awkward could at best mean that one should always *ensure* that one is not clumsy or awkward, and that people who are clumsy or awkward failed to ensure this. Arguably, this is not the kind of criticism people intend to make when they criticize someone for being clumsy or awkward.

Whatever the criticism involved in these thick concepts might be, the criticism of *irrationality* is not a criticism for *bringing* oneself into a state – which is just shown by the fact that we can be allowed to bring ourselves into an irrational state. Thus, if the criticizability of irrationality is to be understood as analogous to the criticizability of clumsiness or awkwardness, this might well mean that someone who is irrational is in some sense criticizable. But premise (3) of the main argument of this subchapter would not be true when we employ this sense of ‘criticizable.’

I do not think that ‘irrational’ is a thick concept,²⁴ and especially I do not wish to rest my argument on one specific account of thick concepts. Instead, I wish to draw our attention to the fact that there are ways of criticizing by which we (a) imply that a person ought not to be the way they are (in a specific sense) and yet (b) they need not have violated any *practical requirement*.

The personal criticism that is involved in ascriptions of irrationality is different from what we might call *moral* criticism. The person subject to rational criticism does not *deserve*

²³ For an account along these lines, cf. Hare 1952/1964, 111-126; 1981, 17-18.

²⁴ Benjamin Kiesewetter suggested to me that ‘irrational’ is a thin concept, like ‘good’ or ‘right.’ This is compatible with my argument I develop here. It is noteworthy that Lord (2018, 4) compares the criticism in ‘irrational’ with the criticism in ‘stupid.’ He thus seems to think that criticizing someone for being irrational is similar to criticizing someone as stupid. But it is doubtful that being stupid is a state that is responsive to reasons, although it might manifest itself in such states. Maybe one normative consequence of regarding someone as irrational can be to regard them as stupid. But I think that this cannot be quite the essence of the criticism involved in regarding someone as irrational.

reactions like resentment, indignation, or even punishment merely in virtue of making a rational mistake.²⁵ Especially when we conceive of cases where a person's environment had a damaging influence on the person's rational disposition, resenting them for an instance of irrationality would often be inappropriate.

This indicates that the criticism involved in irrationality-ascriptions implies a kind of responsibility which is closer to what Gary Watson (1996) called 'attributability.' The kind of blame we are sometimes subject to when an attitude or behavior is attributable to us does not imply the violation of a practical requirement: people might inculpably end up being greedy or arrogant and thus appropriately subject to attributability-blame. But being greedy or arrogant is itself neither an action nor necessarily responsive to reasons, and so being greedy or arrogant need not imply the violation of a requirement, although greedy or arrogant people are appropriate targets of attributability-blame merely in virtue of being greedy or arrogant: they suffer a vice, a defect in character. In this respect, attributability-blame is also different from someone considered as clumsy or awkward. If you consider someone to be clumsy or awkward, then you do not necessarily think that they have a vice.

Both the criticism that does not presuppose a defect in character (clumsy, awkward) and the criticism that does presuppose such a defect (greedy, arrogant) does not amount to the kind of blame we are legitimately subject to when we violate a practical requirement (as when we act morally wrong). This latter form of blame can be equated with Watson's *accountability*-blame, which might come in the form of passionate emotions like resentment, anger, indignation, or guilt. The appropriateness of these latter *passionate* reactions, I will argue at length in chapter 11 of this book, presupposes that the person could have reasonably avoided what we blame them for.

The reactions we have towards someone who is irrational seem to be quite different than the reactions we have towards someone who acts morally wrong, and more similar to the reactions we have towards someone we judge to be just greedy or arrogant: we might think, in some way, badly of this person. But we do not think that the person did something they ought not have done in some *robust* sense: we do not necessarily take them to be *accountable* (in Watson's sense) for their irrationality. This makes it plausible that the violation of a norm of rationality does not have quite the same normative consequences as someone's morally wrongdoing you through their voluntary conduct. In order to understand the criticizability of

²⁵ Similar points were recently made by Brown (2018) and Kauppinen (2018) about our reactions to violations of epistemic standards.

irrationality, we need to find a form of criticism that is (a) different from criticism directed at non-reasons-responsive states (clumsy, awkward), (b) different from our paradigms of moral passionate blaming, and (c) similar to criticism that is directed at character flaws (greedy, clumsy).

Before turning to this form of criticism, there is one objection I need to consider. This is probably the most obvious – but easily rebuttable – objection against the claim that we do not react passionately to other’s irrationality. It seems that we often feel indignant about someone’s irrational denial of human-induced climate change. It thus seems that I am committed to saying that our indignation is not appropriate, for the person did not wrong us by failing to make an effort of will. Rather, they were just irrational. However, it seems implausible to say that our indignation is not appropriate. Thus, contrary to what I claim, there are cases of appropriate passionate blame directed at someone for their irrationality.

In reply, I will distinguish between *culpable* and *non-culpable* irrationality. There is irrationality that we could have reasonably avoided, and irrationality which we could not have reasonably avoided. It seems that most deniers of human-induced climate change have plenty of evidence available to them nowadays that should rationally convince them that they are wrong. Like someone who steals another’s property from minor motives, these people fail to make an effort of will which they owe to others. Their irrationality has severe consequences for other people, and they could avoid their irrationality. This is, I propose, why it is sometimes appropriate to react to them with indignation or even resentment.

Furthermore, indignation might be justified for pragmatic reasons: it might be important to cultivate blaming-responses towards people who are non-culpably ignorant of human-induced climate change in order to, say, promote awareness of the importance of the issue. These responses would then not be appropriate, according to my proposal, but they could nevertheless be justified in an all-things-considered sense (i.e., it would be justified to cause ourselves to inappropriately blame them). Whether this is indeed all-things-considered justified is also an empirical question insofar as the consequences of our blaming the blameless would play a role.

Compare the culpable denier of human-induced climate change with the non-culpable denier. The latter person might be someone who grew up in an environment where human-induced climate change is irrationally denied by everyone. Even if presented with clear evidence which can be understood by all, this group of deniers refuses to acknowledge the scientific conclusions. Being socialized in this group, such a denier might be trained from an early age on to be irrational when it comes to this topic. Considering the evidence even closer and without bias is not anything that comes up as a reasonable course of action for them. Knowing

this social background of the denier, we can say that they lost some authority over their beliefs when it comes to the issue of human-induced climate change.²⁶ This is why resenting them or being indignant would not be appropriate.

Yet their irrationality is a vice, a defect in character. It is attributable to them in virtue of the fact that it is part of their overall outlook on the world, rather than one of people's occasional lapses which we could excuse. In what sense, according to my account, can we still blame the non-culpably irrational deniers of human-induced climate change?

6.4.3 CRITICISM, IRRATIONALITY, AND VICE

To get some idea of the kind of criticism or blame that is appropriate in cases of non-culpable irrationality, consider first how we might make sense of a case of conflict as a case where we cause ourselves to be a way one ought not to be. Take the case of the greedy person:

(G) S ought to cause themselves to become a greedy person.

Even if we could save the world by engaging in a long and successful project of making ourselves greedy, we will still end up being greedy. Our greediness will have some specific normative consequences that are independent of how the greediness came about. By being greedy, we impair our relationship to other people. This normative consequence of an impaired relationship provides other people with reasons to relate differently to us in their actions and attitudes. The way they can appropriately relate differently to us is a form of criticism.

Maybe most obviously, we might not *trust* a greedy person in many contexts. As it is with epistemic distrust (cf. ch. 4.3), we would not want to provide them with information that would lead them to more greedy conduct. Part of being greedy is being epistemically irrational: the greedy person takes themselves to have reasons to believe that a specific situation is a good opportunity to make money, and that they should use this opportunity even if there are strong reasons not to do so, etc. I might not want to tell my greedy friend about a new business deal of which I heard, because this might lead them to re-structure their whole decisions so as to get the most out of this deal for themselves, and to thereby, for example, neglect their duties towards significant others. The greedy person is someone who does not only fail to respond correctly to their reasons for the attitudes that make up their greed, but they are also disposed towards further kinds of irrationality due to their greed. As Owens puts it when discussing

²⁶ In the terminology employed by Fischer and Ravizza (1998), we might say that they lost their *ownership* over the relevant belief-forming mechanisms.

epistemic vices like gullibility, when I am vicious, then ‘I cannot be trusted to think and feel as I ought’ (2000, 124).

More generally, if we come to see a person’s vices, we might reconsider our relationship with them.²⁷ We might want no longer to be friends with the person. We might cease to promote their personal projects or not take pleasure in their successes when we see that they adopt their aims and choose their means only because they are greedy, weak-willed, cowardly, intemperate, ungenerous, unjust, gullible, or dogmatic. We might then also no longer want to please the person, and no longer accept credit from this person for our own actions. We might doubt their judgments more generally because of a general sense of distrust we develop towards them. These are all reactions we can only show to fully responsible beings, because we can only have the relationships that are presupposed by these reactions towards fully responsible beings. Neither computer nor children can display such defects in character that give rise to the reactions described above.

My suggestion, then, is that the kind of criticism we deem to be appropriate when people are irrational, and only because they are irrational, is analogous to the criticism we deem appropriate because people are vicious. This is plausible because many vices involve irrationality. We might even want to say that what distinguishes the criticism involved in calling someone ‘clumsy’ from the criticism involved in calling someone ‘greedy’ is that the latter kind of criticism is grounded in the fact that the greedy person is irrational. If a person was fully rational in all their attitudes and actions, then they would not be *greedy* – even if their conduct might look like the conduct of a greedy person. Rather, they would be in a tragic epistemic position from which it seems to them as if they should behave in a way a greedy person behaves. By behaving that way, they would respond correctly to all the reasons available to them, and thus would escape criticism.²⁸

We thus always *ought* to be rational in a similar sense in which we always *ought* to be virtuous (or at least ought not be irrational in the sense in which we ought not to be vicious). This solves the puzzle presented to us by cases of conflict: for it is both plausible that we can conceive of cases where we ought to make ourselves having a vice, and yet this vice does not

²⁷ The kind of effect such vices can have on our interpersonal relationships is nicely captured by Scanlon’s (2008) account of blame, which is taken up and refined by Hieronymi (2004; 2019) and Smith (2013). This kind of blame will help me in part IV to identify one of the faces of responsibility for attitudes.

²⁸ This last point would require a more extensive argument. I return to this issue in ch. 10.3, where I contend that I do not have a decisive argument for this claim. Yet this will not affect my overall reply to the problem of mental responsibility: I can still say that there is a sense in which we are responsible for our (non-)compliance with reasons for attitudes even if it is not (non-)compliance to one’s *available* reasons that gives rise to criticism or blame.

thereby lose its specific normative consequences: it impairs our relationship to others in specific ways no matter under what conditions it came about. Irrationality can have the very same normative consequences, mainly because it seems that when we blame people because of their vices, we blame them because they are irrational.

How should we evaluate Hussain/Kiesewetter's case (see 6.4.1 above) in light of this discussion? First, I agree that by *calling* the religious people in the example 'irrational' for believing in god, we thereby communicate that we see our epistemic relationship to them as impaired in some way. It might be impaired in that we do not give weight to their testimony when it comes to, say, certain metaphysical questions, and that we do not trust them to process certain types of information correctly (e.g., they might, without sufficient evidence, see certain natural events as 'god's wrath'). As a result of our distrust, we might withhold certain information from them.

Secondly, however, I did not claim that being irrational is *always* criticizable. Remember that the first premise of the arguments from blameworthiness (cf. chapters 4 and 5) only requires that one *can* be blameworthy merely for being irrational. It is thus consistent with my proposal that we deem the religious people's beliefs irrational and yet do not regard them as blameworthy. Aren't we all irrational in certain areas? As long as we do not call out each other for being irrational, we can coherently assume that being irrational is sometimes totally fine, and need not amount to displaying a criticizable vice. *Calling* someone 'irrational' is expressing blame, but merely *judging* someone to be irrational is not necessarily blaming the person. Compare the fact that merely judging someone having *acted wrongly* is not necessarily blaming the person. We are all irrational from time to time, and we all act wrongly from time to time. We can usually excuse each other for occasional lapses and do not regard these lapses as having any significant consequences for our interpersonal relationships. Yet they are lapses nevertheless – i.e., they are genuine cases of irrationality or wrong conduct.

6.5 SUMMARY

In this chapter, I have spelled out another argument that seems to favour nihilism, thereby providing more substance to our challenge of understanding the status of reasons for attitudes. Understanding this status requires us to respond to this *argument from cases of conflict*, and I outlined a reply to it, which will provide the basis for my account of mental responsibility.

The argument states that we sometimes ought to cause ourselves to flout the standards of rationality, and that it is implausible that in such cases that we still ought to be rational (6.1).

The latter claim is (AAT). While evidentialists usually reply to such cases with introducing two-domains of normativity (practical and epistemic), I have argued that the defender of (AAT) can present a powerful argument to back up their claim that puts pressure on evidentialism and – more generally – mental constitutivism. What is more, if this argument is sound, it is just a small step for the mental nihilist to establish their position (6.2).

In reply to the defender of (AAT) and the nihilist, I have suggested that we understand the criticizability invoked in irrationality as analogous to the way we blame people for their vices. This blame need not involve passionate emotions like resentment or indignation – especially when the vice or the irrationality was caused non-culpably. Yet there are ways of relating to vicious or irrational people that presuppose their responsibility, and that can thus be rightly called by the name ‘blame.’ This makes it plausible that in cases of conflict, we cause ourselves to be a way we ought not to be. Yet we should not always regard ourselves as blameworthy in such cases. For if our irrationality does not have any normative consequences for our interpersonal relationships because it was just an occasional lapse, then we are excused for violating standards of rationality.

Since the only premise of the defender of (AAT) that seemed to be questionable in his response to the two-domains reply (cf. 6.2.3) was premise (5), I conclude that my argument must be understood as rejecting this premise:

- (5) If it is false that you ought to [ϕ and ψ], but you ought to ϕ , then it is false that you ought to ψ .

Note that in cases of conflict, ‘ ϕ ing’ and ‘ ψ ing’ refer to two distinct types of responses to reasons – an action of *causing* an attitudinal state, and an attitudinal state. Saying that you ought to [*cause* yourself to be irrational and *be* rational] thus does not make any sense. This is because, as I will point out in more detail below (‘Summary and Conclusion of Part II’), the meaning of the ‘ought’ that is supposed to be at use here is dubious. If failing to perform right actions has normative consequences that are distinct from failing to be rational, then there cannot be one ‘ought’ that requires to give such a mixed response. The two-domains reply was right in pointing out that we have to distinguish the normativity of action from the normativity of the mind. In taking premise (5) as a meaningful claim, the proponent of (AAT) just assumed that there is no such separation between the two realms of normativity. But this assumption was wrong. Looking closely at the forms of blameworthiness or criticizability involved in the violations of practical requirements, on the one hand, and the violations of rational standards, on

the other, reveals that the relevant ‘oughts’ – and thus the reasons – we flout when violating the respective norms have different meaning.

Let me now return to these meanings in more detail, thereby re-evaluating the two central arguments of part II, and outline my solution to the problem of mental responsibility in relation to the discussion up to now.

Summary and Conclusion of Part II

Part II presented us with the challenge of accounting for the normative force of object-given reasons in terms of blameworthiness or criticizability, thereby defending their status as reasons. The challenge came from two arguments that posed the threat of mental nihilism, i.e., the claim that there are no reasons for attitudes. In this closing section of part II, I will re-state the arguments and my reactions to both, thereby also spelling out a reading of the arguments in which I think they are sound: they refute the idea that attitudes can ever be directly required in the way actions can be directly required. We have to acknowledge an ambiguity in ‘ought’ and ‘reason.’

First, our challenge came from the argument from blameworthiness (chapters 4 and 5) that argued with a plausible condition a consideration has to fulfill in order to count as a reason: it need at least be possible to be blameworthy merely for failing to comply with the consideration. I pointed out that I think that non-compliance with epistemic reasons and – more generally – object-given reasons for attitudes *can* make us blameworthy in a sense. This is my reply to the argument from blameworthiness. Yet the way we blame someone merely in virtue of their irrationality will never amount to passionate forms of moral blame – resentment, indignation, or guilt.

Secondly, our challenge came also from the argument from cases of conflict that called into question that object-given considerations ever provide us with reasons for attitudes. In cases of conflict, we ought to cause ourselves to be irrational (i.e., to fail to respond correctly to our object-given reasons). It is puzzling how such cases are possible as long as we assume that the sense in which we *ought* to be rational is the same sense in which we *ought* to perform right actions. I have argued that this assumption is misleading, and thus sketched my reply to the argument from cases of conflict based on a distinction between different forms of blame.

The normative consequences of both kinds of ‘ought’ seems to be quite different. We saw this because our blaming-reactions towards irrationality are more akin to the blaming-reactions we show towards vicious characters than to our blaming-reactions towards morally wrong conduct where another person wrongs someone by failing to make an effort of will they owe to them. This was an initial sketch of how I think we should account for the normative force of object-given reasons: being irrational – i.e., failing to respond correctly to our reasons for attitudes – has the specific normative consequences of impairing relationships in various

ways, similar to the ways in which being vicious impairs our relationships. We can legitimately mark these impairments with various non-passionate reactions.

The distinction of the two ‘oughts’ will allow for readings under which both the argument from blameworthiness as well as the argument from cases of conflict is sound. Yet these readings under which the arguments are sound do not pose any trouble for the existence of mental normativity. They merely pose a problem for the idea that attitudes can ever be directly required in the same way actions can be directly required.

To see this, let us agree on the following definition: ‘S ought_R to ϕ ’ means that one *rationally* ought to ϕ – i.e., it is true that S ought_R to ϕ in virtue of the fact that S’s not ϕ ing can have the normative consequences – and only these consequences – that are associated with failing to be rational (that is, allowing for impaired relationships in specific ways that can legitimately give rise especially to forms of distrust); ‘S ought_V to ϕ ,’ on the other hand, means that one *voluntarily* ought to ϕ – i.e., it is true that S ought_V to ϕ in virtue of the fact that S’s ϕ ing can have the normative consequences associated with failing to make an effort of will that S owes to another person (that is, allowing for wronging someone in a way that can legitimately give rise to resentment, indignation, or guilt). Analogously, we can define a ‘S has a reason_R to ϕ ’ and ‘S has a reason_V to ϕ ’ as saying that one has those kinds of reasons that give rise to the respective kinds of ‘oughts’ if the reasons are decisive.

Note that ‘ought_R’ and ‘reason_R’ is not only applicable to attitudes, but also to wrong conduct: our relationships can be impaired in specific ways not only because of our attitudes, but also because of our actions. Saying that one ought_V to ϕ is thus a very *specific* way of requiring someone to perform an action.

Given these definitions, consider the following two versions of the arguments that make up our challenge:

The argument from mental blameworthiness reloaded

- (1) Considerations of kind K are *reasons_V* for or against an attitude only if we can *legitimately be blamed with passionate emotions* just because we do not have an attitude that seems to be decisively supported by considerations of kind K.
- (2) We cannot *be legitimately blamed with passionate emotions* just because we do not have an attitude that seems to be decisively supported by object-given considerations.
- (3) Thus, object-given considerations are, it turns out, not *reasons_V* for or attitudes.

The argument from cases of conflict reloaded

- (1) S *ought*_V to cause themselves not to be rational. (*cases of conflict are possible*)
- (2) If S *ought*_V to cause themselves not to ϕ , then it is false that S *ought*_V to ϕ . (AAT)
- (3) So, it is false that S *ought*_V to be rational.

Given my definitions, the arguments are sound. Yet their import is not as spectacular as one might hope for. The conclusion of the first argument merely states that object-given reasons are no reasons that can, if decisive, amount to an ‘ought’ the violation of which can give rise to resentment, indignation, or guilt. The conclusion of the second argument merely states that in cases of conflict, S is ought not to be rational in the sense of an ‘ought’ the violation of which can give rise to resentment, indignation, or guilt. I think that much is uncontroversial.

Yet having the distinction between these types of ‘ought’ in mind will ultimately allow us to evaluate the plausibility of *The easy solution*. For the kinds of blameworthiness and criticizability that are made appropriate by violating the different kinds of ‘ought’ plausibly have different preconditions. We are only appropriately subject to resentment, indignation, and guilt if we have wronged someone by failing to make an effort of will. That is, if we direct our passionate emotions at someone because they have, say, an offensive attitude, we thereby assume that the person *could have managed their attitude*: we presuppose indirect control. *The easy solution* thus provides the right preconditions for the type of responsibility that is the basis for the passionate blaming responses. This kind of responsibility in fact presupposes indirect control.

Yet the other kind of responsibility – the responsibility that merely gives rise to the non-passionate blaming responses that mark the impairments of relationships – obviously does *not* presuppose indirect control. For to distrust someone who is vicious is appropriate even though the person did not choose to be vicious. Even if a person was normatively required to make themselves vicious – or irrational – this does not affect the appropriateness of certain forms of distrust and other attitudes that mark the impairment of a relationship with another responsible being.

Understanding the distinction between the two forms of responsibility is the key to evaluating *The easy solution*. Part IV of this book will thus be devoted to defending the distinction between the two kinds of responsibility in more detail. Yet Part IV will not develop a detailed account of the kinds of responses that are made appropriate by irrationality. This is because seeing that there is a difference between the two kinds of responsibility giving rise to these different sets of responses is already sufficient to solve the problem of mental

responsibility. For seeing this difference also makes us see that the preconditions on the two types of responsibility are different. As pointed out before, giving a full account of the normative force of rationality is beyond the scope of this book.

Before turning to this defense of the distinction between two kinds of responsibility, we need to evaluate the other route one might take in reply to the challenge of accounting for the normative force of object-given reasons. Remember that there were three logical ways open to react to the challenge: one can either bite the bullet and deny that there are any reasons for attitudes (*nihilism*); or one can dispute arguments that make up the challenge and argue that object-given considerations for attitudes *have* normative force, and can thus count as reasons (*constitutivism*, i.e., the option I endorse); or one can argue that object-given reasons are not the only type of reasons for attitudes. The latter option amounts to *pragmatism*: the claim that there are practical reasons for attitudes as there are practical reasons for action. If pragmatism was true, then we could just understand the normative force of reasons for attitudes as analogous to the normative force of reasons for action – because the reasons would ultimately be the same and favour actions and attitudes in exactly the same way. I reject pragmatism in the next part III of this book.

Part III

Reasons for Attitudes and Value

Introduction to Part III

We started our investigation by considering the plausibility of *The easy solution* to the problem of mental responsibility. *The easy solution* claims that responsibility for attitudes is completely analogous to the responsibility we have for consequences of our actions. The main worry for this solution was that attitudes are responsive to reasons and thus subject to the norms of rationality – in this respect, attitudes are different than many other consequences of our actions.¹ In part II, we considered arguments to the effect that attitudes are *not* held for reasons in order to see whether *The easy solution* can handle this worry by denying that attitudes are held for reasons. We already saw that these arguments, though intuitively powerful, have their weakness.

I proposed that we should, rather than endorsing these arguments hastily, understand them as confronting us with a *challenge*. The challenge consists in giving an account of the normative force of reasons for attitudes. In order to meet this challenge, we need to understand two features of reasons for attitudes. We need to understand how we are blameworthy or criticizable for failing to comply with decisive reasons for attitudes, and, thereby, in what sense we always *ought* to be rational in believing, desiring, feeling, and intending even in cases in which we ought to *ensure* that we are not.²

Both arguments that presented us with this challenge – the argument from blameworthiness (chapters 4 and 5) and the argument from cases of conflict (chapter 6) – rest on the implicit assumption that whether we comply with reasons *matters* in some way. Given this assumption, we can reconstruct the following general argument as underlying both arguments:

- (1) All reasons for attitudes are object-given.
- (2) (Non-)Compliance with reasons matters.
- (3) (Non-)Compliance with object-given reasons does not matter.

¹ See the Introduction and my discussion of consequential conceptions of doxastic responsibility in ch. 4.4.

² The question of whether we ought to be rational is identical to the question of whether we always ought to comply with decisive reasons for attitudes if we assume a reasons-based conception of rationality (Kiesewetter 2017; Lord 2018), rather than a conception mental coherence (Broome 2013). Following the extensive defenses of Kiesewetter and Lord, I adopted a reasons-based conception of rationality (cf. ch. 0.2.4). If it would turn out that rationality cannot be understood as responding correctly to reasons, then my overall argument is not affected. This is because I am primarily interested in the normative force of (object-given) reasons rather than of coherence. If you do not agree with reasons-based conceptions of rationality, you are free to read my use of 'rational' as stipulative: it refers to the property of an attitude of being adequately based on the subject's accessible reasons for this attitude.

(4) Thus (from (2) and (3)), there are no object-given reasons for attitudes.

(5) Thus (from (1) and (4)), there are no reasons for attitudes.³

According to (1), reasons for belief are solely provided by facts indicating the truth of the believed proposition, i.e., they are provided by the evidence; reasons for desire are provided by facts indicating (or constituting) the goodness of the desire's object; reasons for emotions are provided by facts about the emotion's object which make it rational (or correct) to have the emotion towards that object; and reasons for intention are provided by facts about the intended action that indicate (or constitute) the action's rightness.

I endorse both (1) and (2). If I want to avoid the unsettling conclusion (4), I have to deny (3): Compliance with object-given reasons *does* matter. I have made some initial remarks about how I think it matters.⁴ Yet I did not provide a general account of the significance of the standards of rationality. Since I did not spell out *how* compliance with reasons for attitudes matters yet, the challenge left us in the dark about the sense in which reasons for attitudes, and thus the 'ought' of rationality, have normative force.

At this point, the traditional pragmatist about reasons for belief enters the stage.⁵ They wave their hands in excitement, proclaiming that they have an account of the significance of the norms of rationality which I ignored until now. Traditional pragmatists argue that we should understand the normative force of reasons for attitudes as completely analogous to the normative force of reasons for action. Compliance with reasons for attitudes *matters* in the same sense as it *matters* to comply with practical requirements. In excluding practical reasons for attitudes from the start (cf. ch. 0.2.4), I just misconstrued the nature of reasons for attitudes – according to pragmatists.

Pragmatists think that we should tackle (1): we should deny that reasons for attitudes are all object-given. They argue that I misconstrued the nature of reasons for belief (and, they add, of reasons for other attitudes as well). Reasons for belief are not only provided by facts that indicate the truth of the belief's content – they are not merely object-given. Rather, a type

³ We get the two arguments which were central to part II if we spell out how compliance with reasons is supposed to matter according to the arguments. To get the argument from blameworthiness, we need to substitute 'can make us blameworthy' for 'matters.' To get the argument from cases of conflict, we need to substitute 'can make us blameworthy even in cases in which we ought to cause ourselves not to respond correctly to these reasons.' Premise (1) is an additional assumption that was not mentioned in the arguments from part II because I assumed throughout that part that the only reasons for attitudes there could be are object-given.

⁴ Cf. esp. chs. 4.3, 6.4.

⁵ I will later refer to this traditional version as 'pragmatism.' I will defend a non-traditional version of pragmatism in ch. 7.1 below. The distinction will be relevant for my claim that traditional pragmatism implies that we can form attitudes at will.

of *state-given* reasons concerned with how good it is to have a certain belief (or another attitude) – i.e., *practical reasons* – determine whether it is rational or irrational to believe (or to desire, feel, or intend) something.⁶ Rather than engaging in the hopeless project of understanding the normative force of reasons for attitudes without grounding these reasons in value, we should just grant that reasons for attitudes are grounded in value in the same way as reasons for action are grounded in value: we ought to believe what is best to believe, because we can believe for the same type of reasons for which we can also perform actions.⁷

Thus, pragmatism about reasons for attitudes – i.e., about the norms of rationality – presents itself as an attractive position to endorse: pragmatism seems to be a good candidate for explaining the normative force of reasons for attitudes, for it claims that this normative force is analogous to the normative force of reasons for action. Since I assume the normative force of reasons for action to be unproblematic in this investigation, pragmatism provides a very straightforward answer to our challenge of accounting for this force from part II. Yet, as I will argue in the present part III, pragmatism comes with too big of a burden: since pragmatism implies that we can believe, desire, feel, or intend at will – i.e., that we can believe and form other attitudes for practical reasons –, and since there is no such ability, we should reject pragmatism.

If my argument in part III is sound, then we are left with only two options of how to account for the normative force of reasons for attitudes: either we bite the bullet in response to the seeming insignificance of object-given considerations and grant that these considerations lack normative force, and thereby commit to mental nihilism; or we explain the normative force of reasons for attitudes not by grounding these reasons in practical value, but by arguing that object-given reasons have normative force in themselves, and thereby commit to constitutivism. Since the first option leaves it mysterious why reasons for attitudes are rightly called ‘reasons,’ I endorse the second strategy. Part IV will present the outlines of such a strategy by focusing on the difference between the normative force of reasons for action and reasons for attitudes, and thereby providing a satisfying solution to the problem of mental responsibility.

⁶ Remember that not all state-given considerations are practical in this sense, e.g., the consideration that an attitude is coherent with other attitudes is also a state-given consideration about this attitude (cf. Kolodny 2005, 551). My argument in this part of the book is concerned with practical considerations, i.e., a subtype of state-given considerations.

⁷ I use the phrase ‘what is best to do’ as a placeholder. Spelled out, it means ‘to do whatever your favorite theory of right action says that you ought to do.’ ‘What is best to believe’ thus just means ‘what your favorite theory of right action would recommend to believe if believing was an action (or at least subject to the same kinds of norms than actions are subject to).’ For the sake of simplicity, I sometimes talk as if the true theory of right action was some objective version of maximizing agent-neutral act-consequentialism.

Chapter 7

What Is Pragmatism About Reasons for Attitudes?

Pragmatism is usually presented as the claim that the norms of the rationality of belief are not purely epistemic. Rather, there are practical reasons for belief, i.e., reasons provided by the value of having the belief. It is thus sometimes rational to believe something because it is valuable to harbour the belief. This is true, according to pragmatists, because we can, at least sometimes, believe for practical reasons.¹ Evidentialists deny that we have such an ability, and that, because of this, there are no practical reasons for belief.² If there are no practical reasons, then we should understand the norms of belief as purely epistemic.³

Analogously, *mental pragmatism* would claim that the norms of the rationality of attitudes are *not* purely derived from the constitutive aims of attitudes (i.e., truth, desirability, rightness, etc.). Rather, the fact that an attitude is valuable to have can be a practical reason for this attitude. This is because we can, at least sometimes, form or adopt or maintain attitudes for practical reasons. *Mental constitutivism* – the generalized analogue to evidentialism – denies this ability. It claims that since we do not have any such ability, the norms of rationality that govern our attitudes are to be solely derived from the constitutive aims of attitudes.

¹ We can distinguish between *moderate pragmatism* and *radical pragmatism*. McCormick (2015) and Reisner (2008; 2009) are moderate pragmatists: they argue that at least *some* reasons for belief are practical. Rinard (2015; 2017) is radical: she argues that evidence does not provide us with reasons *at all*, and that thus *only* practical reasons are reasons for belief. Radical pragmatism would provide a more straightforward account of the normative force of reasons for belief: compliance with these reasons matters in the same way as compliance with reasons for action matters, because the reasons are just of the same kind. Moderate pragmatists need an account of how epistemic and practical reasons play together in determining what we ought to believe (cf. McCormick 2017; Reisner 2008). They thus need to say something about the normative force of epistemic reasons as well. It thus seems that moderate pragmatists do not have any argumentative advantage compared to constitutivism: both constitutivism and moderate pragmatism face the challenge of explaining the normative force of epistemic reasons. However, I will just grant to moderate pragmatists here that they might have a plausible way of explaining the normative force of epistemic reasons which is not an option for me – for example, they could endorse epistemic instrumentalism (cf. ch. 4.2.3). I thus will not treat the fact that moderate pragmatists also need to say something about the normative force of epistemic reasons as an objection against their view. Rather, I argue directly against practical reasons for belief.

² Cf. chs. 3.2, 3.3. The line of argument sketched in this sentence is the one brought forward by Shah (2006) which I already mentioned in ch. 3.3, and which I discuss also below (ch. 7.2).

³ Evidentialists usually do not deny pragmatic factors to influence what is rational to believe. For example, according to what is discussed as ‘pragmatic encroachment’ in epistemology (cf. the volume of Kim/McGrath 2019), the evidential threshold that is required to epistemically justify a belief depends on the practical stakes at play if one is wrong. These practical stakes, however, do not enter as reasons *for which we believe*. Thus, even if we grant pragmatic encroachment, the norms of belief can still be purely epistemic.

We will see in this chapter that these ways of formulating doxastic or mental pragmatism are not fully clear. This is because the notion of ‘believing for practical reasons’ or, respectively, ‘forming an attitude for practical reasons’⁴ is ambiguous. I argue that there is a trivial sense in which we *can* believe for practical reasons: we can believe *indirectly* for practical reasons. We will see how this merely allows for an uncontroversial form of pragmatism which I will call *non-traditional* pragmatism. Yet there is another, far from trivial, sense of ‘believing for practical reasons.’ To say that we can believe for practical reasons in this sense is to say that we can *directly* believe for practical reasons – i.e., believe at will.

I will argue in this chapter that only the latter, controversial version of pragmatism – traditional pragmatism – is in conflict with evidentialism and constitutivism, but not the former, uncontroversial non-traditional pragmatism which I endorse. We should thus understand pragmatism, insofar as we want to understand it as a controversial claim, as implying that we can sometimes believe at will. The next chapter will then argue that there is no such ability, and thus refute the controversial traditional pragmatism, and thereby establish mental constitutivism.

I start with defending my version of *non-traditional* pragmatism (7.1), before I turn to my argument that in order to defend a controversial version of pragmatism – i.e., a version that contradicts evidentialism and constitutivism –, one needs to argue that we can believe at will (7.2). This sets the ground for my argument against pragmatism in chapters 8 and 9, where I will argue that pragmatism fails because we cannot believe (as well as desire, feel, intend) at will.

7.1 Ought to believe and reasons for action

We can see the uncontroversial form of pragmatism if we take pragmatism as an analysis of ordinary statements of the form ‘S ought to believe that p’ or, respectively, ‘S ought to have a mental state M.’ The uncontroversial form of pragmatism then claims that these sentences are not merely made true by epistemic reasons. Rather, they imply practical reasons in some way. Spelling out this ‘in some way’ will make us see how this version of pragmatism is uncontroversial and does not conflict with evidentialism. Having this version of pragmatism clearly in

⁴ In what follows, I will most times merely mention believing as the relevant response. For most of my discussion, it does not matter about which kind of attitudinal response to reasons we talk. If we cannot believe at will, we can also not feel at will or desire at will. If doxastic pragmatism fails, then pragmatism with respect to most other attitudes fails as well. This is because my argument against believing at will in chs. 8 and 9 will build on the ontology of attitudes *in general*. I will talk separately about intentions and decisions in ch. 9.3.

front of our mind will be important for my argument that traditional pragmatism implies believing at will.

Following Wilfried Sellars, Matthew Chrisman (2008) argues that ‘rules of criticism’ or ‘Ought-to-be’s,’ like ‘Clock chimes ought to strike on the quarter hour,’ imply ‘Ought-to-do’s’ (‘Somebody ought to bring it about that clock chimes strike on the quarter hour’). It would not make any sense to say that ‘Clock chimes ought to strike on the quarter hour,’ with this way of employing ‘ought,’ if there was either no one who is able to set the clock in the right way or if there was no one who has a reason to do so.

Analogously, Chrisman argues, sentences like ‘One ought to believe in human-induced climate change’ can be analyzed as saying that we should *ensure* that we (or others, like our children)⁵ believe in it. This use of ‘ought to believe’ is frequent in the philosophical literature. As I pointed out already in my discussion of Clifford and James (chapter 1), Clifford discusses doxastic duties under the heading of ‘The Duty of *Inquiry*’ (1877, 70, my italics), and William James emphasizes how we can purposefully *allow* our passionate nature to influence our beliefs (1896, 11). Both authors seem to have discussed the ethics of belief as concerned with how to *deal* with our beliefs, rather than as concerned with the norms of rational belief itself.

Let me give two more examples of this use of ‘ought to believe’ within philosophy. Henry H. Price writes ‘that 19th century ladies acknowledged a moral obligation *to believe* that their husbands or fiancés were impeccably virtuous’ (1954, 13, my italics). Yet Price ultimately understands their (purported) ‘duties to believe’ in terms of how these women should *control their attention* in order to cultivate beliefs that show their husbands in beneficial lights. Thus, Price does not balk one moment at the idea that ‘S ought to believe’ is to be analyzed in terms of reasons for action rather than reasons for belief.

Similarly, McCormick notes, referring to Stroud (2006), that the demands of friendship are such that your beliefs about your friend should be somewhat unresponsive to the evidence. Stroud suggests that you should interpret what you hear in a less-damaging way than would a stranger, looking for alternative interpretations to the obvious, and damning, ones. Just as a certain way of *behaving* concurs with our ideas of what a good friend would do, so, she argues, does a corresponding manner of *believing* (McCormick 2015, 60)

It is important to see here that the ‘manner of *believing*,’ at least according to this description, *consists* in a manner of *behaving* – for interpreting is obviously some kind of intentional mental

⁵ For the social dimension of Chrisman’s account of epistemic normativity, cf. Chrisman *fc.*

activity. The idea that our doxastic duties depend on our reasons for certain actions seems very natural to Clifford, James, Price and McCormick – and I think it does to most of us.

‘Duties to believe’ in this sense are uncontroversial in that they are just duties to act in certain ways: they are not duties that are in place independent of voluntary control; rather, having indirect control over our beliefs is a precondition for such duties to exist. As a result, what I called consequential conceptions of doxastic responsibility (chapter 4.4) are suited to make sense of intuitions of pragmatists: in a sense, what we ‘ought to believe’ is a matter of practical reasons – more specifically, a matter of practical reasons for action, rather than for belief.⁶ In this sense of ‘ought to believe,’ we can say that we ought to believe what is best to believe in the same sense as we ought to do what is best to do (the ‘ought’ to believe just is an ‘ought’ to do). This sense is just the sense in which we should *ensure* that we have the best beliefs, which is not a controversial claim if we understand it properly.

It is a commonplace among pragmatists that our beliefs should sometimes be somewhat unresponsive to the evidence when a lot is at stake. There is a sense in which this might be true about advantageous but epistemically not fully rational beliefs about our own capabilities,⁷ about significant others,⁸ or about religious matters.⁹ That we ought to *cultivate* such an unresponsiveness to the evidence can be *expressed* by saying ‘We ought to *believe* that p’ when ‘p’ refers to an epistemically irrational belief about ourselves, about others, or about god’s existence as long as the ‘doxastic’ Ought is understood as a duty that is ultimately just a duty to *cultivate* good beliefs.¹⁰

Whether this position makes sense with respect to other attitudes will depend on whether there are uses of ‘S ought to feel angry,’ ‘S ought to be happy,’ ‘S ought to desire X,’ ‘S ought to intend to ϕ ’ that imply reasons for action. I think such uses are very common when it comes to negative emotions with bad consequences: ‘do not get angry at him, he is not worth it,’ or ‘do not allow yourself to be pulled down by this tragic event.’ By giving such emotional

⁶ Remember that I have argued in 4.4 that consequential conceptions of doxastic responsibility, and thus *The easy solution*, cannot explain epistemic normativity. They might, however, provide a plausible account of indirect responsibility for belief and thereby underwrite pragmatic duties of belief-management. However, I have argued that this responsibility does not exhaust our responsibility for belief – there is also a direct responsibility to epistemic norms.

⁷ Hazlett (2013, 44-52) cites psychological studies which suggest that overestimating one’s own capabilities slightly promotes one’s wellbeing.

⁸ See McCormick (2015, 60-61), who follows Stroud (2006). However, see also Crawford’s (fc) argument that the demands of close interpersonal relationships cannot be such that they require unresponsiveness to the evidence.

⁹ See McCormick (2015, 61-65). The classical point of reference for the idea that we should believe in god even in the absence of sufficient evidence, see Pascal (1670). See also James (1897).

¹⁰ I defended such a duty in Schmidt (2017).

advice, we tell the other person to re-focus their attention away from an object that would provide them with decisive object-given reasons for anger or, respectively, sadness. Focusing our attention on certain aspects of our situation is, arguably, sometimes a helpful technique for managing one's emotions. Also the Stoic advice not to get unsettled by seemingly horrible events (like the death of a significant other) might be read as an advice to manage your emotional life in a beneficial way and independently of whether the event is *in fact* horrible.¹¹

For my systematic purposes, it does not matter whether we have a use of 'S ought to have M' with respect to all types of attitude M such that this use implies reasons for action. Keeping in mind the possibility that there might be attitudes where we do not have any such established use of 'S ought to have attitude M,' we can formulate non-traditional pragmatism as follows:

Non-traditional pragmatism: Reasons for action are among the truth-conditions of a type of sentences of the form 'S ought to believe, desire, feel, or intend, (etc.) that'

Non-traditional pragmatism is not an exhaustive account of mental normativity. It is not even an account of reasons *for attitudes*. Rather, it is a reductive analysis of a certain use of 'ought to believe, desire, feel, or intend': it claims that, properly understood, sentences of this form sometimes just state that we ought to bring it about, or ensure, or maintain that we have a certain attitude. What looks like a genuine attitudinal 'ought' on the surface is ultimately just an 'ought' to perform certain actions. Non-traditional pragmatism thus reduces *one aspect* of mental normativity to the familiar normativity of action, thereby making this aspect of mental normativity intelligible.

It is important to keep this uncontroversial form of pragmatism in mind. For remembering it will allow us to see that recent arguments that are intended to establish pragmatism merely establish non-traditional pragmatism, rather than its more controversial cousin – traditional pragmatism – to which I return below.

What is the lesson for evidentialism and constitutivism to draw from this? Evidentialists will not be impressed by non-traditional pragmatism. They will grant that they did not talk

¹¹ I think of Epiktetus' advice to tell oneself, if a precious cup breaks, that 'it was just a cup,' and if one's wife dies 'it was just a human' (Epiktet, *Encheiridion*, art. 3). We might want to doubt that the Stoic strategy of dealing with life's tragedies is psychologically healthy. However, also the Pyrrhonists might be understood as advising us to cultivate a mind that is beneficial to our *ataraxia* – i.e., a mind free from definite assumptions about what is in fact good or bad. Arguably, neither the Stoic's nor the Pyrrhonist's advices are advices to comply with our object-given reasons. Rather, they aim at having attitudes that contribute to a mind that is *peaceful*. On an interesting treatment of Pyrrhonism as an art of living, cf. Olfert *fc*.

about duties to believe that are *reducible* to duties to act. Rather, they were talking about *irreducible* duties to believe: they were talking about the norms of rational belief, not of rational (or right) action of attitude-management. Thus, in order for evidentialism or constitutivism to be false, another version of pragmatism would have to be true:

Traditional pragmatism: The norms of rationality that govern our attitudes are not to be (solely) derived from the constitutive aims of attitudes (truth, desirability, rightness, etc.). This is true because there are (also) practical reasons for attitudes (in addition to object-given reasons), i.e., we can form attitudes for practical reasons.

In order to conflict with evidentialism, ‘forming attitudes for practical reasons’ has to be understood as responding *directly* to a practical consideration by forming the attitude. If we would merely indirectly respond to a practical reason – i.e., by performing an action for this reason – , then this would not amount to the kind of ability presupposed by traditional pragmatism. The ability to respond *indirectly* to practical reasons merely underwrites non-traditional pragmatism.

I will now consider some recent arguments of pragmatists and argue that they merely establish non-traditional pragmatism. My discussion will allow me to develop an argument that concludes that traditional pragmatism implies that we can form attitudes *directly* for practical reasons. Chapters 8 and 9 then deny this ability.

7.2 TRADITIONAL PRAGMATISM IMPLIES MENTAL VOLUNTARISM

I did not provide a full account of what we *irreducibly* ought to believe, desire, feel, or intend, even though I sketched what I take to be the meaning of the relevant ‘ought’ in chapters 4.3, 5.2, and 6.4. Yet by aiming to provide a solution to the problem of mental responsibility by highlighting the difference between requirements of action and standards of rationality, I committed myself to presenting the outlines of such an account of the normative force of the standards of rationality. Traditional pragmatism (from now on: pragmatism) is an account of what we *irreducibly* ought to believe, desire, feel, or intend that grounds the normative force of this ‘ought’ and the corresponding reasons in value (analogously to the way many practical reasons are grounded in value). If pragmatism fails, I must explain the normative force of reasons for attitudes without grounding these reasons in value in this way. I turn to this task in part IV after we saw why pragmatism fails.

The present chapter argues that pragmatism is not a plausible account of what we irreducibly ought to believe, desire, feel, or intend. This is because

- (1) pragmatism implies mental voluntarism, i.e., it implies the claim that we can believe, desire, feel, or intend directly for practical reasons, and
- (2) mental voluntarism is false.

Especially McCormick, Reisner and Rinard have recently argued for pragmatism by arguing that we can believe for practical reasons in the same way as we can act for practical reasons (McCormick 2015; 2017b; Reisner 2008; 2009; Rinard 2015; 2017).¹² Both Rinard and McCormick present their arguments for pragmatism in response to Nishi Shah's (2006) argument for evidentialism. I will thus first present his argument briefly (7.2.1) before I turn to their responses (7.2.2, 7.2.3). I develop my argument that pragmatism implies voluntarism by discussing their responses. I finally summarize my argument (7.2.4).

7.2.1 SHAH'S ARGUMENT FOR EVIDENTIALISM AND Φ ING FOR A REASON

Shah's (2006) argument for evidentialism can, a bit simplified, be re-stated as follows:

- (1) R can only be a reason to ϕ if one can ϕ *for* this reason.
- (2) We cannot believe for non-evidential reasons.
- (3) Thus, only evidence can be a reason for belief.¹³

I think that, interpreted in a certain way, this argument is just fine. I can accept the conclusion because I do not think that there are other reasons besides those provided by evidence that can be properly called 'reasons for belief.' The reasons on which I rely in my non-traditional pragmatism are reasons for actions, rather than practical reasons for belief (see chapter 7.1 above).

Premise (1) seems to be a plausible restriction on what can count as a reason for something. Shah spells out the notion of ' ϕ -ing for a reason' in terms of the function a reason can play in deliberation about whether to ϕ (Shah 2006, 485). When we think about whether to

¹² I will not discuss Reisner here because he does not propose any arguments for believing for practical reasons that are different in essence from the argument which Rinard spells out in more detail than he does (for his short argument why indirect control over belief is sufficient for believing for practical reasons, see Reisner 2009, 68-70).

¹³ The argument for evidentialism in Kelly (2002) can also be summarized with this syllogism, for he relies on the premise that consideration can only be a reason to ϕ if ϕ ing can be *based* on this consideration. The basing-relation he has in mind is the same as the one Shah has in mind when he talks about the relation of ϕ ing for a reason.

believe that p , it seems that we focus solely on the evidence. This phenomenon was labeled ‘transparency’ by Richard Moran (1988): the question of whether to believe that p is transparent to the question of whether p . It thus seems that practical considerations cannot enter into our deliberation about whether to believe that p , because practical considerations are not concerned with whether p , but rather with whether it would be good to believe that p . Premise (1) thus allows for a test of whether something deserves its status as a reason by considering whether it can play a specific role in our reasoning. According to this analysis of ‘ ϕ ing for a reason,’ that you ϕ for a reason means that you base your attitude on a reason in such a way that we could plausibly conceive of your attitude as the conclusion of your deliberation or reasoning about whether to have that attitude.¹⁴

That we cannot believe *for* practical reasons thus means that practical considerations cannot enter into our deliberations about what to believe, which would allow us to conclude, together with (2), that practical considerations are not reasons for belief. According to Shah’s argument, practical considerations may enter as *causes* into the formation of our beliefs, but never as *reasons*: the thought that it might be good to believe that p may cause me to acquire the belief, but the acquired belief would then not count as having been acquired *for a reason*. I return to the difference between a belief’s being caused in reaction to a practical consideration and believing for practical reasons in chapter 8.2 below.

I will sometimes use the term ‘practical *consideration*’ instead of ‘practical *reason*’ because I ultimately agree with Shah’s argument: there are no practical *reasons* for belief, even though there are practical *considerations*, i.e., contents of thought that present a belief as valuable to have. I will only speak of ‘practical *reasons* for belief (or for an attitude)’ for the purpose of presenting the pragmatist’s point of view. Practical considerations are not necessarily

¹⁴ More passive attitude-forming processes (like subconscious drawing of inferences, or perception) should not be an obstacle for this analysis of ‘ ϕ ing for a reason.’ This is because even if we form an attitude without explicitly engaging in deliberation, we can retrospectively make sense of the fact that you have this attitude by constructing a syllogism that makes it intelligible to us why you have this attitude. E.g., if you form the desire to eat X without explicit deliberation, we could, retrospectively, ascribe the following subconscious reasoning process to you: (1) I should eat something healthy now; (2) X is something healthy I could eat now; (3) thus, I should eat X . (3) makes it rational for you to desire to eat X , and the fact that we can ascribe the syllogism (1)-(3) to you as a subconscious reasoning process makes it thus rationally intelligible to us why you desire to eat X (i.e., it makes it intelligible to us under the general assumption that people are rational).

One might ask: What is the syllogism we can construct for basic perceptual beliefs, like the belief that I am reading right now? It is not clear to me that we can make brute perceptual beliefs rationally intelligible, or if we should rather just say that they are not held for reasons (but rather just caused by the environment). This is a deeper epistemological issue which would lead us to discussions about foundationalist and coherentist accounts of epistemic justification, and to the status of Wittgensteinian ‘certainties’ (*Gewissheiten*). I need not get into this. However, if you think that perceptual beliefs are based on reasons, I think you are committed to the possibility of constructing a plausible syllogism that makes it rationally intelligible that people with perceptual beliefs hold the belief. Such a syllogism might involve facts about how things appear to you in perception.

reasons for anything – either because they are not always facts (for I assume that all reasons are facts) or because we cannot always react to them.¹⁵

What could one object against Shah's argument? One could either object that (1) is too strict a condition on being a reason to ϕ ; or one can argue that we can in fact believe *for* practical reasons, and thus deny (2). Both Rinard (2015) and McCormick (2015) argue that, given a certain reading of 'believing for a reason,' (1) is false, but given another reading of 'believing for a reason,' (1) is true; but that if we adopt the latter reading, then (2) is false.

I think both arguments contain an important insight. Yet their arguments do not show that there can be other reasons for belief than reasons provided by evidence. The practical reasons for belief Rinard and McCormick argue for are just reasons for actions that influence our beliefs, rather than reasons for belief. Their arguments thus support non-traditional pragmatism, rather than the controversial traditional pragmatism that would conflict with Shah's evidentialism. I will examine both accounts in turn, and thereby show that traditional pragmatism implies that we can believe at will – i.e., that we can believe *directly* for practical reasons.

7.2.2 WAYS OF Φ ING FOR A REASON

Rinard argues that evidentialists need to understand (1) as the claim that a consideration R is a reason to ϕ only if we can ϕ *directly* for R, and that (1) is false under this interpretation. She argues that (1), interpreted in this way, is too strict a condition on what can count as a reason to ϕ . This is because (1) would also rule out reasons for certain actions. Her example illustrates this argument:

[S]uppose you're getting ready to go snowshoeing, and you are trying to decide which socks to wear. Last time you wore cotton socks, and your feet were too cold. You know wool is warmer than cotton, and so you'll be more comfortable if you wear wool. Surely this constitutes a genuine pragmatic reason in favor of wearing wool socks. But note that appreciating this consideration does not lead you *directly* to wearing them; it leads you *directly* only to an intention to take steps toward wearing them, such as opening your dresser drawer, finding and removing the wool socks, etc. So it fails Shah's requirement for a consideration to count as a genuine reason. But surely the fact that you'll

¹⁵ Considerations that are false (i.e., no facts) but that would still favour a response if they were true (i.e., if they were facts and thus reasons for the favoured response) can be called *apparent* reasons (Alvarez 2018, 3003-4). Cf. ch. 0.2.2 on the concept of reasons. Considerations to which we *cannot* react are not even apparent reasons. For example, take the consideration that believing in god would be pleasant. Let us assume both that we cannot believe at will and that there is no course of action available to me (of which I know) that would with some likelihood result in my believing in god. In this case, the consideration that believing in god would be pleasant is not a reason for me to do anything because I cannot react to it.

be most comfortable if you wear wool socks *does* constitute a genuine reason for wearing them (Rinard 2015, 212-213).

Note first that, according to Rinard, we can distinguish between ϕ ing *directly* for a reason and ϕ ing *indirectly* for a reason. It seems uncontroversial that I can believe indirectly for a practical consideration: I can sometimes *bring about* a belief through an action for the reason that it is good to have the belief (think of Pascal's recommendation of inducing belief in god by participating in the religious life). Shah and I think that this indirect control over belief is not sufficient to transform this practical consideration into a *reason for which we believe*. Rather, we would say, the consideration is merely a *reason for which we bring a belief about*. Rinard denies this: if we bring about a belief because we think it would be good to have the belief, then we can be said to believe for a practical reason.

Rinard supports her claim that we can believe for practical reasons by arguing that the relation between a response and a reason in ' ϕ ing for a practical reason' is the same when ' ϕ ing' is a complex action (like wearing wool socks) and when ' ϕ ing' is believing. She points out that the fact that I have to first form the intention to wear wool socks, and then open the drawer, get them out, and put them on before I finally *wear* them is not in conflict with the fact that I end up wearing the wool the socks *for a reason* – namely, the same reason for which I formed the intention to wear them, opened the drawer, etc. Similarly, the fact that I first have to form the intention to bring a belief that p about and then take certain steps to bring the belief that p about before I finally *believe* that p is not in conflict with the fact that I believe *for a reason* – namely, the for the same reason for which I brought my believing about, which was a practical reason. Thus, Rinard concludes, there are practical reasons for belief.¹⁶

To evaluate this argument, we first need to distinguish between actions and mere states of affairs that are not responsive to reasons. We can agree that, for example, your *putting on wool socks* is an action you do for a reason. *That you are wearing wool socks* is a state of affairs. You cannot bring about *that you wear wool socks* directly as a response to reasons, for this state is not a response to reasons. Saying that you have a reason to be in the state of wearing wool socks can at most mean that you have reason to *ensure* that you are in that state. Compare: 'You should be outside longer – it is such a nice weather today!' This sentence does not imply that you can *just like that* be in the state of 'being outside' by responding to the stated reason, even though, in some sense, the sentence states a *reason to be outside*. Saying that someone

¹⁶ For similar, but less elaborated arguments for believing at will, cf. Stocker 1982, 410 and Reisner 2009, 68-70.

should be in a state or has a reason to be in a state can only be elliptical, however. A ‘reason to be outside’ is a reason to *ensure* that you are outside.¹⁷ Thus, if Rinard’s point is that we can be in states for reasons in the same way as we can act for reasons, then her claim is false. At most, what we can do for reasons is to *ensure* that we are in states (because ensuring is an action).¹⁸

However, Rinard’s other examples suggest that this was not her point. For sure, she could reply, to be in the *state* of wearing wool socks is nothing we do for reasons in the same way as we perform basic actions for reasons. However, she could add, *to wear wool socks for a reason* – i.e., to perform a *long-term activity* for a reason – involves the same relation as performing a basic action for a reason. Rinard gives more examples of long-term activities done for a reason:

For example, that air pollution is worse in the city is a reason to live in the country; that it’ll be easier for you to pay attention during a talk if you sit in the front row than the back is a reason to sit in the front; that doing so would be relaxing is a reason to spend some time in Costa Rica; etc. But none of these considerations plays the particular role in the regulation of one’s ϕ -ing identified by the Evidentialists as necessary for it to count as a genuine reason (Rinard 2015, 213).

Rinard’s claim is that we cannot just live in the country, sit in the front row, or spend time in Costa Rica as a *direct* response to a reason, as we can perform basic actions as direct responses to reasons or form beliefs as direct responses to the evidence. Nevertheless, the reasons for which we perform those activities are reasons *for* those activities.

Rinard’s activities are *long-term activities* which aim at bringing about or keeping up a certain state of affairs – the states that you live in the country, sit in the front row, or spend

¹⁷ The ‘should’ might also sometimes be merely *evaluative*, i.e., merely state that it is good to be outside. However, it need not. It might very well be meant to imply that one is able to bring the good state of affairs about and that one has decisive reasons to do so (cf. ch. 7.1 above).

¹⁸ One could object that we never directly respond to reasons by acting, because our only direct responses to reasons are attitudes. When we act, it is the intention to act that is a direct response to the reason, rather than the action itself. It seems that Rinard endorses such a view when she claims that only intentions can directly regulate our actions, but that reasons for action cannot (2015, 215). But if this was Rinard’s point, then she would claim that we can *never* act directly for reasons. This is surely odd, and, as Raz points out, it ‘is to misconceive the relations of intention and action, imagining that when acting intentionally one acts by forming an intention that causes one to act’ (Raz 2011, 57). Even if one is not willing to accept that one can put on wool socks as an immediate response to reasons, one would have to accept that one can perform the basic actions involved in putting on wool socks (moving one’s limbs) for reasons. It might even be true that we mean different relations when we talk about ‘acting for a reason’ on the one hand, and ‘believing for a reason’ on the other (I think Rinard commits herself to this when she rejects the idea that ϕ ing for a reason requires forming the intention to ϕ , and then to ϕ (2015, 215)). This, however, does not make it the case that talk of ‘being in a (non-reasons responsive) state for a reason’ (or, as I argue below, our talk of ‘performing a long-term activity for a reason’) involves the same relation as believing for a reason.

time in Costa Rica. To defend premise (1) of Shah's argument, we could reply that, strictly speaking, we are not performing the long-term activities for reasons.¹⁹ What we do for reasons are rather the more basic actions which make up the long-term activities. That is, if we say that we have a reason to live in the country, what we mean is that we have a reason to perform the more basic actions which make up our living in the country. But, strictly speaking, there are no reasons for 'living in the country.'

This claim would be controversial, and so I do not wish to endorse it. Instead, I think that the fact that we cannot perform long-term activities as direct responses to reasons justifies introducing different senses of 'φ-ing for a reason.' Rinard states that in the case of long-term activities 'the causal connection between the pragmatic consideration for φ-ing, and the agent's actually φ-ing, is complex and indirect. But this does not prevent the consideration from constituting a genuine reason for φ-ing' (ibid.). This, however, highlights the fact that there is an important difference between performing long-term activities for a reason, on the one hand, and performing more basic actions for a reason or believing for a reason, on the other. This difference can be captured by introducing different senses of 'φ-ing for a reason.' In the first sense, 'φ-ing for a reason' requires that one can φ directly for this reason in the same way as one can directly perform a basic action, like lifting one's arm, for a practical reason, or believe directly on the basis of one's evidence.²⁰ In the second sense, we can also 'φ for a reason' if we do not φ *directly* for this reason – for example, when performing long-term activities, where the relation between the reason and the activity is 'complex and indirect.' Thus, there is a sense in which we can live in the country for a reason, but this sense is different from the one in which we can believe, or perform basic actions, for a reason.

What Rinard's examples thus show is that we can believe for practical reasons in the same sense in which we can perform long-term activities for practical reasons, or be in (non-reasons-responsive) states for practical reasons. It thus follows that (1) can be given two interpretations, and Shah's argument for evidentialism is sound only according to the first one:

(1') R can only be a reason to φ if one can φ *directly* for this reason.

(1'') R can only be a reason to φ if one can φ for this reason *by performing basic actions which either constitute the φing (long-term activity) or result in φing (state).*

¹⁹ Rinard considers and rejects this objection (2015, 214-215).

²⁰ As mentioned in fn. 18 above, one may wish to make another distinction within this first sense if one thinks that the relation in 'acting directly for a reason' is different than the one in 'believing directly for a reason.' The distinction which is relevant for the following argument, however, is not this one.

If we choose the reading of ‘ ϕ ing for a reason’ employed in (1"), then premise (2) is false under this reading of ‘ ϕ ing for a reason,’ for one obviously *can* believe for practical reasons in this sense. Thus, in a sense, Rinard’s objection to Shah’s argument is sound. However, and most importantly, that we can believe for practical reasons in this sense of ‘ ϕ ing for a reason’ is *not* in conflict with evidentialism. That we can believe indirectly for practical reasons merely establishes the uncontroversial *non-traditional* version of pragmatism according to which statements of the form ‘S ought to believe that p’ (or, for that matter, statements of the form ‘S has a reason to believe that p’) are statements about *what belief to bring about* or *what belief to maintain* (or, respectively, statements about reasons for bringing about or maintaining a state).

Thus, all Rinard can show with her argument is that we can ‘believe for practical reasons’ in the sense that we can be in states for reasons – i.e., we can have reasons to bring ourselves into states or maintain these states, including belief-states.²¹ Rinard’s insight establishes that some sentences of the form ‘S ought to believe that p’ *mean* that there are decisive reasons for actions of bringing the belief about, or for actions of maintaining the belief. This does not show that *irreducible* reasons for belief (i.e., reasons for belief *proper*, which are not reducible to reasons for action) can be practical reasons. Rinard’s argument does not contradict evidentialism, and thus not establish traditional pragmatism. What evidentialists wish to deny is that we can believe *directly* for practical reasons in the way we can perform basic actions *directly* for practical reasons, or in the way we can believe *directly* on basis of our evidence.

7.2.3 DELIBERATION AND Φ ING FOR A REASON

McCormick objects to Shah’s argument that we have to understand doxastic deliberation more broadly so as to encompass practical reasons as objects of doxastic deliberation (McCormick 2015, 28-30). According to her objection, Shah’s argument only works if ‘believing for a reason’ requires that the consideration for which we believe can be a premise in a syllogism which concludes in the content of the belief for which the consideration is then a reason. If by ‘deliberating’ we mean mentally operating with such syllogisms, then no practical reasons can enter into deliberation, because practical reasons do not (normally) speak in favour of p’s truth.

²¹ One could object that Rinard’s argument leaves open the possibility that we can ‘believe for a reason’ in the same sense in which we can perform long-term activities for a reason, which is a different sense of ‘ ϕ ing for a reason’ as the sense involved in ‘being in a state for a reason.’ In response, I will explicitly argue against the claim that believing is an activity in chapter 9. I do not wish to deny, however, that believing might be essentially *accompanied* by long-term activities, like by what Chrisman calls ‘maintaining a coherent system of beliefs’ (Chrisman 2016) – an activity that might be crucial for understanding indirect doxastic norms and forms of doxastic responsibility, but which is distinct from believing itself.

According to McCormick, however, this is too strict a condition on what can count as a reason for something. She points out that there is a broader understating of deliberation: by ‘deliberating about what to believe’ we can mean something that is more akin to practical deliberation. In practical deliberation, we might mentally make a ‘pro- and contra-list’ with respect to whether perform some action. Now, if such a list is concerned with whether to adopt a belief, then practical reasons for belief can be part of this list, according to McCormick. Premise (2) would thus turn out to be false if we understand ‘believing for a reason’ in terms of this more permissive picture of doxastic deliberation.

Again, I think that the insight in this argument does not refute evidentialism, for evidentialists can again restrict their position to the first meaning of ‘believing for a reason.’ Like Rinard, McCormick introduces two meanings of the phrase, which can be captured as follows:

- (1*) R can only be a reason to ϕ if one can ϕ for this reason, *and where ‘ ϕ ing for this reason’ requires that the reason can enter as a premise into a syllogism which concludes in ϕ ing.*
- (1**) R can only be a reason to ϕ if one can ϕ for this reason, *and where ‘ ϕ ing for this reason’ merely requires that the reason can be part of pro- and contra-list which is concerned with whether to ϕ .*

Given the second meaning of ‘ ϕ ing for a reason,’ however, it is again not clear in what sense ϕ ing can count as an *immediate* response to the reason. Indeed, McCormick seems to fall prey to the same problem as Rinard. The strict understanding of ‘ ϕ ing for a reason’ in (1*) would make Shah’s argument sound and exclude practical reasons for belief. But the more permissive understanding in (1**) does not show that we can believe *directly* for practical reasons. Thus, the only thing it can show is that there is a sense of ‘believing for a reason’ which is compatible with the fact that we cannot believe directly for this reason. McCormick’s argument thus does not support traditional pragmatism, but rather merely the more trivial non-traditional pragmatism.

The arguments of both McCormick and Rinard point to meaningful doxastic ought-sentences which are made true by reasons for actions. But they do not show that conceptual evidentialism is false, as long as conceptual evidentialists restrict themselves to those doxastic ought-sentences that do not imply reasons for actions.

7.2.4 PRAGMATISM IMPLIES VOLUNTARISM: Φ ING FOR PRACTICAL REASONS IS Φ ING AT WILL

Both Rinard and McCormick implicitly change the topic by interpreting Shah's premise (1) in a sense in which nobody denies that we can believe for practical reasons. I will thus ignore this uncontroversial use of ' ϕ ing for a reason' as employed in their premises which I labeled (1'') and (1**) in each of the two previous sections. What is interesting is whether we can believe for practical reasons in the sense employed in (1') and (1*), because this would mean that we can believe for practical reasons as a direct response to these reasons in the same way as we can act for practical reasons as a direct response to these reasons – it would establish mental voluntarism.²²

Let me finally summarize my argument why (traditional) pragmatism implies voluntarism. This is established by showing that ϕ ing for practical reasons, understood as the capacity we need to have in order for traditional pragmatism to be true, is ϕ ing at will:

- (1) To ϕ for a practical reason either means that you ϕ indirectly for this reason or it means that you ϕ directly for this reason.
- (2) To ϕ for a practical reason does not mean that you ϕ indirectly for this reason.
- (3) Thus, to ϕ for a practical reason means that you ϕ directly for this reason.

I established (2) in my arguments against Rinard and McCormick. If you indirectly ϕ for a reason, you merely do something for a reason that leads you to ϕ ing (if ' ϕ ing' means to be in a state) or that partly constitutes ϕ ing (if ' ϕ ing' means that you perform long-term activities). Indirectly ϕ ing for a reason is not the sense of ' ϕ ing for a reason' in which we can perform basic actions for a reason, or believe for an epistemic reason. Yet this is the sense of ' ϕ ing for a reason' pragmatists need if they want to make a different claim from non-traditional pragmatism. Thus, to ϕ for a practical reason – in the sense *traditional* pragmatists need to use this phrase to establish their position – means to ϕ *directly* for this reason.

I did not explicitly argue for (1). One could try to deny (1) on the grounds that the distinction between ϕ ing directly for a reason and ϕ ing indirectly for a reason is not logically complete distinction: if you ϕ for a reason, then you can ϕ for this reason in a third way that is neither direct nor indirect. However, it is hard to see what the third option could be. If you ϕ for a reason, then you must ϕ either as an immediate reaction to this reason ('to ϕ directly'), or

²² Reisner's argument that indirect control over belief is sufficient for making believing for practical reasons possible (Reisner 2009, 68-70) provides us with the same insight. It does not show that we can believe for practical reasons in the sense in which we can believe for epistemic reasons. It just shows that we can do so in the sense of (1'').

your reaction is mediated by another reaction ('to ϕ indirectly'). I.e., your reaction is either not mediated or it is mediated. There is no third option. Thus, (1) states a logically complete distinction and is thus analytically true.

Finally, one might wonder why I equate believing at will with ϕ ing for *practical* reasons. That is, one could doubt that it follows from (3) that to ϕ for practical reasons means to ϕ at will. Of course, one might directly ϕ for *other* kinds of reasons – e.g., for object-given reasons, as when one directly believes that p for the reason that some other fact indicated that p . This is not believing at will. If you believe at will, you do not base your belief on your evidence, but rather on facts about how valuable it would be to have the belief. The idea is that ϕ ing at will is to exercise the kind of control that we also exercise when we act for reasons.²³ Since to act for reasons means to act as a direct response to *practical* reasons, ϕ ing at will means to ϕ as a direct response to practical reasons. It would not help the pragmatist at all if believing at will would mean to believe directly in response to *epistemic* reasons, because the claim that we can do so is uncontroversial and does not support any form of pragmatism. Since the claim that we can believe at will is meant to support pragmatism in the traditional version, it is best understood as the claim that we can believe directly for practical reasons.²⁴ I now turn to my argument why we lack such an ability.

7.3 SUMMARY

I started out in this chapter by distinguishing non-traditional pragmatism from traditional pragmatism. I argued that non-traditional pragmatism, which is a reductive analysis of statements about what attitudes we ought to have in terms of reasons for action, is uncontroversial and does not conflict with evidentialism or constitutivism (7.1). I then argued that the recent arguments of pragmatists that pragmatism does not imply voluntarism fail insofar as they are meant to support *traditional* pragmatism (7.2). They fail because their arguments that there is a sense of ' ϕ ing for practical reasons' in which we can believe or form other attitudes for practical reasons merely establishes *non-traditional* pragmatism. These arguments do not establish traditional pragmatism, which is the controversial and more interesting claim. ' Φ ing for a practical reasons,' in the more interesting sense, means that you ϕ *directly* for this reason – i.e.,

²³ More precisely: When we perform basic actions for reasons. For I have argued in 7.2.2 that we cannot perform long-term activities for reasons in the same sense in which we can perform basic actions for reasons: I cannot live in the country at will.

²⁴ Cf. Bennett 1990, 90.

believe, desire, feel, or intend at will. I now turn to my argument that forming attitudes at will is conceptually impossible, and that this is why pragmatism (in the interesting version) fails.

Chapter 8**Forming Attitudes for Practical Reasons**

I will prepare the ground for my argument against voluntarism by first distinguishing between merely reacting to a reason (or a consideration) and reacting *for* a reason – an, as I will argue, uncontroversial distinction that has important implications for two common arguments in the debate about voluntarism. I then turn to my argument against voluntarism. According to the argument, voluntarists face an inescapable dilemma: In order to believe for practical reasons in the way we can act for practical reasons, forming a belief has to be an active and immediate response to the practical reasons in the same way as acting is an active and immediate response to practical reasons; however, any belief-formation that is active is not immediate, and any belief-formation that is immediate is not active. Believing for practical reason thus rests on a conceptually unfulfillable condition.

My argument will have a leak, however. The leak is that the voluntarist can just reformulate their position in such a way that it is no longer affected by the argument I spell out in this chapter, because my argument relies on a specific phrasing of voluntarism: that we can ‘form’ attitudes for practical reasons. I point out this leak in 8.4, where I discuss objections to my argument against voluntarism. I will stuff this leak in chapter 9 by engaging in some *mental ontology*: I argue that the objection that opens the leak only arises when we assume that attitudes are actions – but they are not. Rather, they belong to a more *static* ontological category. Otherwise attitudes could not help *explaining* our actions.

The upshot of part IV is that there is no understanding of ‘believing, desiring, feeling, or intending for practical reasons’ in which our attitudinal response to reasons is both active and immediate in the way performing an action for a practical reason is both active and immediate. Because of this, attitudes are not subject to practical norms like actions are: attitudes cannot be based on practical reasons in the way actions can be based on practical reasons. If this is true, then there is no hope in understanding the normative force of reasons for attitudes in terms of value. We need a different account of how reasons for attitudes matter.

8.1 REACTING TO REASONS AND REACTING FOR REASONS¹

In a first step, I distinguish between merely reacting to a practical consideration and reacting to it *in the right way* so that the reaction can count as believing for a practical reason (8.1.1). The distinction is motivated by cases of deviant causation. The distinction enables me to point out the restricted scope of the most common arguments for and against voluntarism (the argument from beneficial, but epistemically irrational belief; and the argument from ‘the aim of correctness’) and to thereby motivate my own argument against voluntarism (8.1.2-8.1.3). The next section then presents my argument (8.2). The final section considers objections (8.3). One of these objections will require me to say more about the ontology of attitudes – a task I postpone to chapter 9.

8.1.1 REACTING TO, REACTING FOR, AND DEVIANT CAUSATION

As various authors point out, merely believing something as an immediate reaction to a practical consideration is not sufficient for believing for a reason (rather than just for believing causally *because* of this consideration).² Owens notes at one point that ‘[i]n principle, anything can cause anything’ (2000, 102). That my thought that believing *p* would be good *causes* me to believe that *p* does not show that the content of this thought is a reason for which I believe. The thought that it would be good for me to believe in god *may* cause me to believe in god. Maybe it is just a matter of character how often one of our beliefs comes about as an immediate reaction to practical considerations. Nevertheless, cases in which my thought about the value of having a belief *causes* a belief do not yet show that I can believe at will.

Thus, believing as an immediate reaction to practical considerations is not always believing at will. We would have to react to the practical consideration *in the right way* in order to count as active in forming the belief. For voluntarism to be true, the relation between the practical reason and the attitude would have to be the same as the one between a reason and an action performed for this reason. The argument I spell out below (8.3) shows that this relation cannot be the same.

The problem of deviant causation – traditionally seen as a problem for causal theories of action explanation – can be utilized for illustrating the distinction between believing as a

¹ I developed a proto version of the main argument against voluntarism which I spell out in sections 8.2 and 8.3 in Schmidt 2016 (German article).

² Cf., e.g., Bennett 1990, 106-7; Hieronymi 2006, 48; Scott-Kakures 1994, 81. How can considerations, which are just contents of thought, be causes? They can be said to be causes derivatively insofar as the thoughts of which they are contents are causes.

mere reaction to a practical consideration and believing for a practical reason.³ Consider a case where someone shows a behavior as a mere reaction to a consideration:

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally (Davidson 1973, 79).

Nor, we can add for our purposes, does the causation by the belief-desire pair suffice for the climber's behavior to count as an action done for a reason. That the climber wants to let the other man die might just lead to an involuntary reaction in his hands that causally leads to a loosening of his hold. Similarly, that I want to believe in god because having this belief would be pleasant might just causally lead to some psycho-physical changes in my body that result in my believing in god. In such a case, my belief in god is a *mere reaction* to a practical consideration. For I do not believe *for* the practical reason that believing in god is pleasant (even if we assume that it is true that believing in god is pleasant) – as the climber does not react *for* the reason that he needs to rid himself of the weight and danger. Rather, we both merely react to a consideration.

Note that I do not claim that non-deviant causation is necessary for someone to ϕ for a reason. My claim is much weaker. I merely say that the mere causation of a reaction by a thought about a consideration is not sufficient for ϕ ing *for* the consideration the thought is about. Causation of an attitude by a thought about a practical consideration is, at most, *merely reacting* to this consideration.⁴

This uncontroversial distinction between merely reacting to a reason (or a consideration) and reacting *for* this reason (or consideration) that is commonplace in theory of action has two implications for the debate about voluntarism that have been overlooked – it reveals the restrictions of the most common argument for voluntarism as well as the restrictions of the most common argument against it.

³ Cf. Schmidt 2016, 575-576.

⁴ I say 'at most' because there might be causations by thoughts which are not related to the contents of these thoughts at all. For example, my having a thought might make me unaware of my surroundings, and thus cause me to stumble. My stumbling was not caused by the thought in virtue of the thought having the content it had – it could just have been another thought that would have had the same effect. If a reaction was caused by a thought in virtue of the thought having the content it has, we might say, for the sake of brevity, that it was 'caused by the consideration' (as I sometimes say). Mere causation by a consideration is not reacting *for* it, but merely reacting *to* it.

8.1.2 WHAT VOLUNTARISTS MUST, AND MUST NOT, ARGUE FOR

First, voluntarists often point to cases where it is *psychologically plausible* that people form a belief as a reaction to a practical consideration. For example, Carl Ginet (2001, 64) describes a case in which his thought that it would be too big of a hassle for him to turn back and check if he closed his door leads him to believe that he closed his door. Other familiar cases where considerations about the value of a certain belief might plausibly *cause* this belief concern cases where such practical considerations cause optimistic beliefs about one's own survival, or about the virtues of the near and dear, or about religious matters, or about 'framework' beliefs.⁵ When our relationship to a loved one or our own life is at stake, we sometimes tend to believe what we want to believe, rather than what is supported by our evidence. Psychologists are prone to point this out to us. Does rejecting pragmatism mean that we reject what these psychologists say?

Surely not. It should be clear by now that by merely describing such cases we cannot show that the person believed at will: the mere fact that their belief resulted from their practical consideration is not sufficient for them to believe for a practical reason. McCormick (2015, 29) overlooks the possibility of deviant causation when she claims that if practical considerations can enter as *causes* in belief formation, it is hard to see why they should not be able to enter also as *reasons*. However, there can be plenty of causes of our beliefs which are no reasons for them.

The task of the voluntarist is thus to show that we can believe *for* a practical reason in the same sense as we can act *for* a practical reason: in order to control beliefs in the same way as actions done for reasons ('at will'), the relation between the belief and the reason for which we believe needs to be the same as the relation between the action and the reason for which we act.⁶

⁵ McCormick discusses all three kinds of cases in support of her pragmatism (cf. 2015, 53-65).

⁶ To be fair, neither Ginet nor McCormick merely describe such cases, but provide theoretical background. I discuss McCormick's background in chapter 9. Ginet's main support for the idea that in some cases we can believe directly for practical reasons is his dispositional analysis of beliefs as conditional intentions (cf. Ginet 2001, 67): to believe that *p* means to behave in a certain way in certain circumstances. Ginet thinks that it is plausible to suppose that we can form such a disposition for practical reasons. However, it should be clear by Kavka's (1983) toxin puzzle that it is as problematic to claim that intentions are subject to the will as to claim that beliefs are subject to the will: I cannot just *intend* at midnight to drink a mild toxin the next day if I would get a lot of money merely for *intending* it as long as I do not have any reason to actually *drink* the toxin the next day.

8.1.3 WHY AIMING AT CORRECTNESS DOES NOT REFUTE VOLUNTARISM

Secondly, having in mind that causation of an attitude by a practical consideration is not sufficient for voluntarism, we can see that any argument against voluntarism that builds on the idea that attitudes conceptually ‘aim at correctness’ (more commonly: ‘belief aims at truth’) has a quite restricted force. Such arguments can at most establish that it is conceptually impossible to acquire *some* beliefs or other attitudes for practical reasons. This is because these arguments only bear on the question on what kinds of beliefs can be *caused* in us, but not on the question of whether we can believe *for* practical reasons.⁷

There are certain conceptual restrictions on *which* beliefs can be caused in a person. If a desired belief does not fit at all in the rest of my system of beliefs, like the belief that the moon is made of cheese, it may be indeed conceptually impossible that I acquire this belief by causing it. If the content of my beliefs is conceptually restricted by the rest of my beliefs, then I cannot believe something that contradicts too much of my other beliefs, or is at odds with my past and present behavior. In order to be able to acquire a belief immediately, it somehow has to fit in my belief-system, the belief has to be a ‘live option,’ to use the phrase of William James (1896, 2-3). It would only be possible to adopt a belief that is at odds with one’s current belief-system if one thereby acquires a great deal of other beliefs which form a coherent whole, and in which the newly acquired belief fits in. For example, it would be implausible to suppose that an ordinary person of our society could *just like that* acquire a belief that the earth is flat. This belief would, in order to count as genuine *belief* and to remain stable over time, have to come with other beliefs, like beliefs about a huge conspiracy going on. Any futuristic device that could cause such a belief in us would thus need to induce *a whole system* of beliefs.

I take it that this is the main import of Bernard Williams’s (1970) claim that ‘belief aims at truth.’⁸ It is important to see here, however, that this does not yet explain why we cannot believe for practical reasons. It only explains why we cannot just cause ourselves to have *any* belief we like, whether or not it fits with the rest of our beliefs. It might still be the case that

⁷ For the origin of arguments against doxastic voluntarism Williams’ (1970) much interpreted argument that we cannot believe at will because belief ‘aims at truth.’ See also the argument by Winters (1979). For helpful discussions about the sense in which beliefs aim and truth, see Owens (2003) and Steglich-Petersen (2008), as well as the volume by Chan (2013). McCormick (2018) helpfully distinguishes between conceptual arguments against doxastic voluntarism that build on the nature of belief as involving a truth-aim as endorsed, for example, by Adler (2002) and Levy (2007), and arguments that build on the nature of belief as a state. My argument (below in ch. 8.2) can be counted to the second kind of arguments. Among these, McCormick includes Audi (2001), Chrisman (2008; 2016), and Setiya (2013).

⁸ Cf. also Winters’ (1979) claim that a belief that is not backed by evidence cannot be upheld. For a discussion and defense of this claim, cf. Lindner *fc*.

we can adopt *some* beliefs for practical reasons, or adopt a whole belief-system for practical reasons.

Similar points will hold for other attitudes insofar as their nature prohibits certain kinds of incoherence or irrationality.⁹ The more I desire both *p* and not-*p*, the more doubtful it is that I desire any of the two. The more often I fail to intend what I believe I ought to do, the more one might doubt in the future that I intend to do what I claim that I intend to do in the first place. The more I intend to do both *A* and non-*A*, the more it becomes doubtful that I have any of both intentions at all.¹⁰ Yet the fact that our attitudes are subject to these norms of coherence or rationality does not imply that we can never form an attitude directly for a practical reason. These norms just imply that we cannot form any attitude that is totally at odds with the rest of our mental household for a practical reason.

8.2 THE VOLUNTARIST'S DILEMMA

Rather than relying on attitudes' aims of correctness, I propose an argument against voluntarism that makes us aware of ontological differences between actions and attitudes. Let me first give an intuitive idea of the argument before I spell it out in more detail.

To get a first grip on my argument, consider the following dilemma that threatens the strong doxastic voluntarist: Either we react *immediately* to a practical consideration by forming an attitude as an immediate response to this consideration. But then it does not seem to be an *active* reaction (i.e., not a reaction *for* a practical reason). Believing as a mere causal reaction to a practical consideration cannot be believing *for* a reason – it is immediate but not active. Or we react quite *actively* to a practical consideration by *bringing* ourselves to have an attitude that we desire to have. But then we obviously do not bring about the attitude *immediately*. In both cases, we are not forming an attitude for a practical reason. Believing for a practical consideration would have to be an *active and direct response* to this consideration in the way acting for a reason is an active and direct response. Yet it seems impossible, given the two horns of the dilemma, that a belief is an active *and* direct response to a practical consideration.

⁹ I also talk of 'incoherence' even though I accept a reasons-based conception of rationality (cf. 0.2.4), because for the pragmatist, the incoherences I describe would not be forms of irrationality: we do not fail to respond correctly to our practical reasons for belief by exhibiting these incoherences.

¹⁰ Maybe emotions are different here, for the following statement is false: the more I believe that *X* is not dangerous, the more doubtful it is that I actually fear *X*. It might be obvious by looking at a phobic confronted with a spider that he is afraid, even though we know that he knows that the spider is harmless. His belief does not give us any reason to doubt his fear. This might be because we know of the *recalcitrance* of emotions: they seem to be more prone to persist in the face of our better judgments than other attitudes.

Even if I realize certain regularities in myself like ‘whenever I think that it would be good to believe *p*, I usually come to believe *p*,’ and if I can use these regularities by causing valuable beliefs in myself, this does not amount to believing for practical reasons. Rather, in such a case, I might have reasons for thinking about advantages of certain beliefs, or for attending to those advantages, since these activities may cause the advantageous belief in me. This, however, is only *indirectly* influencing my belief via mental activity, not believing directly, or at will.¹¹

The dilemma above is not restricted to *believing*, since it does not rely on anything special about belief. Rather, it holds for attitudes in general: I am not active and in direct control of an attitude just because I form it as an immediate reaction to a consideration which shows it good to have the attitude. The voluntarist thus must show that in some of those cases where we form an attitude as an immediate reaction to a practical consideration, this is not just something that *happens* without our active participation, but rather a controlled (re)action, i.e., something done *for* this reason. But, as I will argue now, this cannot be done.

There is an important ambiguity in the notion of ‘forming a belief,’ noted by Meylan in a footnote (2013, 11, footnote 8).¹² ‘Forming a belief’ can either refer to an activity previous to the moment where the belief comes into existence, or it can refer to the event of the belief’s coming into existence itself. To illustrate the first sense: We sometimes conduct activities like investigating into a question (by, e.g., reading a paper) or by thinking about an issue. Usually, these activities aim at forming a belief about whether *p*. If they do so, ‘forming a belief’ merely refers to a goal-directed activity like an investigation. Since we can investigate into whether *p* or think about whether *p* for practical reasons (some issues are more worth investigating into than others), we can form beliefs for practical reasons in this sense. This sense of ‘forming’ a belief, however, is not what allows us to conclude that we have direct control over our beliefs.

¹¹ Bennett (1990, 93) imagines the folk of the ‘Credamites’ who can systematically cause beliefs in themselves in a very immediate way and who forget afterwards how they came to have the belief (in this way they can maintain their belief even if they originally adopted their belief for practical reasons). Bennett points out that the Credamites are a case where people *induce* a belief. I think this talk about inducement is telling. Bennett struggles for an explanation of why we cannot *induce* beliefs immediately, and comes up with the idea that there is always an ontological mediation between the inducement and the belief by the physical structures on which the belief supervenes (his idea is that we can only change our mind by first changing our body). He rightly rejects this explanation as an explanation of why belief is involuntary, because the immediacy we deem to be impossible is not ontological, as his explanation assumes, but motivational (105-106). My informed diagnosis is that Bennett developed the wrong kind of explanation because he asked the wrong question: the question is not whether we can *induce* beliefs immediately (for here belief would be *motivationally* mediated by an action of *inducing*), but rather whether we can believe directly for practical reasons as we can act directly for practical reasons. Maybe Bennett found the latter suggestion so obviously absurd that he ended up discussing whether we can induce beliefs without ontological mediation.

¹² Meylan calls it ‘acquiring a belief’ – I take the notions to be equivalent.

This is because ‘forming a belief’ in this sense only amounts to controlling our beliefs by actions prior to the formation. If we control a belief by forming it in this way, we bring a belief *actively* about, but we do not bring it *directly* about.¹³

So far, things seem to be clear. We may, however, refer to the event that *results* from such activities as ‘formation of a belief.’ This event is the change of the state of affairs from the state *S does not believe that p* to the state *S does believe that p*. It can only be this second sense of ‘forming a belief’ that is at issue in the debate about doxastic voluntarism. For in the first sense, nobody denies that we can form a belief at will – this is just to say that we can perform certain ordinary activities which can result in beliefs.¹⁴

It is not conceivable what it would be like to *actively* form a belief (second sense) without previous activities. Compare for illustration the idea of *actively finding X*. To find something implies that the world changes from the state *S does not have X* to the state *S has X*. We can only bring about this state *actively* by looking for X. It is not conceivable what it would be like to *actively find* something without looking for it – i.e., finding something ‘directly.’ I may find X directly in this sense when X just lay accidentally in front of me. But in this case, I did not find it actively and I am not responsible for finding it. In the same way, if the formation of a belief just takes place without me influencing it before, there is no sense in which I controlled the formation. Conversely, if I found X actively, this could mean that I found X because I was looking for it. But in this case, I did not find X directly. In the same way, if I influence a belief formation by prior activity, I did not form it directly.

One might object that there is a sense in which ‘to find something’ can refer to an action that is directly controlled (cf. Stocker 1982, 410). Take our scientist, Lara, from chapter 2.1. She might be excited about being close to finding the cure for cancer. As a result of her excitement, she might move around through her laboratory very quickly, documenting her results, and re-checking some instruments to verify her results. We might observe her in awe and describe what she is doing (with eyes wide open): ‘She is finding a cure for cancer!’ Here, the objector claims, we refer to ‘finding’ to the various activities she is performing.

¹³ Note that this was overlooked by authors who want to support voluntarism by arguing that we can form a believe ‘immediately’ in a *temporal* sense, where the belief follows immediately after performing a basic action (cf., e.g., Levy/Mandelbaum 2014). Also Bennett’s thought experiment of the ‘Credamites’ (cf. fn. 11) merely describes beings which can respond to practical reasons *by performing a basic action* which, in a *temporal* sense, issues immediately in a belief. This is just indirect control over belief, and I think it is not up to debate that indirect control over our beliefs is *conceptually* possible – *pace* Bennett’s argument about *ontological* mediation, which I need not evaluate here. As I pointed out in 8.1.3, there *are* some conceptual restrictions as to *which* beliefs we can cause in us.

¹⁴ My example of belief *formation* is just an illustration – it is easy to apply my thoughts also to the *relinquishment* or *maintenance* of belief. There is the same ambiguity in meaning between activity and result.

In reply, I would like to remind of the ontological distinction between changes of events and activities I made above. ‘To find something’ might have the two uses described by the objector. It might, first, refer to the event of change from the state *S does not have X* to the state *S has X* (as described above) – an event which is a punctual occurrence that results from certain activities. Or it might, as the objector claims, refer to the activities from which the finding (in the first sense) results. However, according to the latter use of ‘to find something,’ the analogy to the passive sense of ‘belief-formation’ which I described above does not hold. Rather, finding something in this second sense as a telic activity is analogous to activities like reading a news-paper in order to form beliefs about whether *p*. That there are such activities does not show that we can believe directly for practical reasons, but rather that we perform activities directly for practical reasons that result in beliefs.¹⁵

My argument can be summarized as follows:

- (1) You can only believe (immediately) for practical reasons (i.e., at will) if you either form, relinquish, or maintain a belief for such reasons.
- (2) ‘Forming,’ ‘relinquishing,’ or ‘maintaining’ a belief *either* means (a) that you perform certain actions which lead to a belief (or disbelief), *or* it means (b) that an event takes place which may result, for example, from such activities.
- (3) In the case of (a), you influence the belief by prior activity, i.e., you form it actively, but not directly.
- (4) In the case of (b), the event issues directly in a belief, but is not itself active.
- (5) Thus, there is no case in which you can believe at will.

The argument does, if sound, not only show that *believing* for practical reasons is impossible. When substituting other mental states for ‘believing,’ the argument also shows that feeling, desiring, or intending, or forming other attitudes for practical reasons is impossible. This substitution is unproblematic since the argument does not rely on any features of belief that it does not share with other mental states.

8.3 RE-FORMULATING VOLUNTARISM?

One could object to premise (2) that the two senses of ‘forming a belief’ are not exhaustive. There might be a third sense of forming a belief in which the formation is both active and in

¹⁵ My reply is inspired by Nottelmann’s reply to Stocker (cf. Nottelmann 2006, 565-567). However, Nottelmann just denies that ‘finding *X*’ can refer to an action, while I grant that it might refer to an action but point out why this does not imply that we can form attitudes at will (cf. also my argument in Schmidt 2016, 577-578).

which we control the belief directly. For example, various authors have suggested that we might control our beliefs by judging something to be true.¹⁶ Analogously, we might control our desires by judging things to be good and our intentions by judging certain courses of actions to be right. Maybe we can also sometimes, by judging an attitude to be valuable, directly acquire the attitude for the reason that it is valuable?

Let us grant to the objector that we can form attitudes by performing some kind of activity called ‘judging.’ Even if judging is not an action, it would still be an activity that *results* in an attitude-formation in the second sense distinguished above. Thus, no matter how we understand the idea that we *form* attitudes – whether by performing ordinary intentional actions, or by judging things to be a certain way –, we will never control attitudes as directly as we control an action. We will always control it *via* some other kind of activity.

That philosophers tend to construct activities like judgments in order to make sense of mental agency just illustrates that we cannot conceive of attitudes as direct *and* active responses to reasons. When we think about how we control our attitudes, we tend to construct activities *by means of which* we control those states – formations, acquisitions, judgments, or settlings of questions.¹⁷ By contrast, if we think about how we control our own actions, we notice that there is no need to think of such further activities. It even seems to be misguided to talk of ‘controlling’ our actions – for to control a thing is to control it with something else. We do not have direct control *over* our actions: for what is it by means of which we control our own actions? Rather, we should say that our actions are themselves *exercises of control*.¹⁸

¹⁶ McHugh and Smith rely on the notion of judgment in making doxastic control intelligible. Owens uses the notion of a ‘practical judgment’ in order to explain our responsibility for intentions and actions. However, according to Owens, actions are not only controlled by practical judgments, but also by the will (in contrast to intentions: the will is not itself controlled by the will).

¹⁷ For the latter activity, cf. Hieronymi 2006. In her more recent work (especially Hieronymi ms), she is explicitly stating that settling a question *is* believing, desiring, feeling, intending, etc., rather than an activity prior to it. In her earlier work, the ontological status of settlings of questions seemed more ambiguous. Her claim that attitudes are active states might be understood merely as a way of emphasizing that attitudes are held for reasons (and are thus different from passive sensations, like headaches). However, I am not sure about the philosophical utility of introducing this new sense of ‘activity.’ Hieronymi argues that we need to conceive of attitudes as active states in order to make human agency intelligible (cf. Hieronymi ms). Also Boyle (2013) argues that we need to acknowledge active states in our ontology. These proposals come with their own problems, but they are not intended to support voluntarism. Rather, they want to make sense of our agency we exercise in response to *object-given* reasons.

¹⁸ We sometimes control our own actions when we learn new skills – e.g., when I learn to ride a bike, I might carefully control my movements so that I stay in balance. In these cases, we might be said to control our bodily actions with our mental actions. However, the better skilled we get in what we do, the less we need to *control* our actions by such mental navigation. Rather than *controlling* our own actions, we usually *exercise control* in acting.

This disanalogy with respect to our agency when it comes to actions, on the one hand, and attitudes, on the other, is best explained by the fact that there is a fundamental ontological difference between both kinds of things: actions are the kinds of things to which we are most intimately related in our active nature – the fact that we can perform actions just *is* the fact that we can influence our environment, and, insofar as we know the laws our environment is subjected to, that we can *control* our environment. We bear no such active relation to our attitudes, for otherwise these attitudes would be actions themselves.

However, voluntarists might want to doubt this disanalogy. Even though philosophers tend to construct further activities by means of which we control our attitudes when we think about mental agency, we need not construct such activities. There is no need to formulate voluntarism in terms of exercising our agency by means of ‘formations’ or ‘acquisitions’ of attitudes. Rather, we can just believe, desire, feel, or intend at will. Premise (1) of my argument is false: believing at will is not *forming a believe* at will, it is just *believing* at will.

With this new formulation of their position at hand, the voluntarist can propose the following *reductio ad absurdum* of the voluntarist’s dilemma. Consider the change from the state *that I am not running* to the state of *that I am running*. Can I bring about this change of state *actively* and *directly*? After all, I can only actively bring it about *by running*, that is, by performing an action. This is, by definition, bringing the change about indirectly. It thus seems that I cannot bring about an action of mine *actively and directly*.

Note that talk of ‘bringing about an action’ only makes sense if we conceive of my action as a state: the state *that I am running* (or the state *of my running*). Arguably, my action is not this state. However, the voluntarist can then similarly argue that *that I am believing that p* is not my believing that p. In the same way as I can bring about the state *that I am running* simply *by running*, I can also bring about *that I am believing that p* simply *by believing p*. The voluntarist is thus right in pointing out that premise (1) crucially relies on the assumption that believing that p is being in a state, rather than being active. If believing was a state, then the only way of conceiving of doxastic control is by conceiving of activities by means of which we get into that state (bringings about, formations, acquisitions, or judgments). That philosophers tend to conceive of doxastic agency in terms of these activities just reflects the prejudice that attitudes are states – so the voluntarist. If we drop the assumption that believing is a state, then we might be active in believing in a similar way as we are active in acting.

I take this objection to be the most promising line of thought against my argument. What the mental voluntarist must argue for is that we can *believe* for practical reasons – it is true that they need not argue that we can *form beliefs* for practical reasons. However, this raises

the question of whether believing (as well as desiring, feeling, and intending) is a state. If it is a state, then the only way to conceive of believing (desiring, etc.) at will is by bringing it about, forming it, acquiring it, etc. The voluntarist's dilemma in section 8.3 shows that these forms of control are not sufficient for believing at will, i.e., they are not sufficient for believing for practical reasons: we would only perform the bringings about, the formations, or the acquisitions for practical reasons, and our attitudes would result from these ordinary actions as any state can result from our actions. The voluntarist thus has to deny that attitudes are just ordinary states.

The voluntarist has two strategies. They can either argue that attitude-responses (believings, desirings, etc.) are themselves actions, or that we can believe, desire, feel, or intend for practical reasons even though these attitude-responses belong to a different ontological category than actions. I will discuss these two strategies in turn in the next chapter.

8.4 SUMMARY

By distinguishing between reacting to reasons and reacting for reasons, I have argued that two common arguments within the debate about voluntarism have restricted force: neither the voluntarist's cases about beneficial beliefs that are insufficiently supported by object-given reasons nor the involuntarist's conceptual argument that 'attitudes aim at correctness' can decide the question of whether we can form attitudes at will (8.1). I then proposed a new argument against the possibility of forming attitudes at will (8.2): the voluntarist's dilemma. According to this argument, the voluntarist has to claim that we can both actively and directly form an attitude for practical reasons. Yet the only senses of 'attitude-formation' there are allow merely for direct, but passive attitude-formation, or for active, but indirect attitude-formation. I finally rejected the objection that there is active and direct attitude-formation in response to object-given reasons by arguing that this claim would not support voluntarism, and I pointed out that voluntarist's can try to avoid the voluntarist's dilemma by re-formulating their position (8.3). I will now turn to this re-formulated version of voluntarism and argue that it fails as well.

Chapter 9

Attitudes as States

Chapter 8 concluded that we cannot *form* attitudes for practical reasons. Though mental voluntarism is often presented as the claim that we can *form* attitudes for practical reasons, one might as well argue for a slightly different claim: that we can just *believe, desire, feel, or intend* for such reasons (rather than *form* these attitudes for such reasons). My argument from chapter 8 does not straightforwardly apply to this second version of mental voluntarism, because it rests on a crucial ambiguity of ‘forming an attitude’ – a notion that is irrelevant to the second version of voluntarism. The second version of mental voluntarism comes in two sub-versions: (a) we can believe, desire, feel, or intend at will because attitudes are themselves actions; or (b) we can believe, desire, feel, or intend at will even though attitudes are *not* themselves actions.

I begin by addressing the first sub-version (a) and argue that attitudes are not actions by arguing that they are non-dynamic, and thus do not fit any active category to which actions belong (9.1). An objection to the effect that attitudes need to be conceived somehow as ‘active states’ insofar as they are responsive to reasons gives me some opportunity to sketch some elements of my picture of mental agency (9.2), and especially to address the worry that our intentions or decisions are somehow more active than other attitudes (9.3). I finally discuss the second version of mental voluntarism (b) by arguing that attitudes and actions cannot both be controlled in the same way if they are ontologically different kinds of things (9.4). My discussion there will be guided by a close examination of McCormick’s (2015) account of doxastic control as guidance control which I interpret as arguing for the second sub-version of mental voluntarism (b).

9.1 ATTITUDES AND ACTIVITIES

One might wonder whether my understanding of attitudes is far too passive. After all, attitudes are responsive to reasons. They are not like headaches or mere physical features of the body. One way to argue that we can believe, desire, feel, or intend for practical reasons is to exploit this idea and push it a bit further: if beliefs, desires, intentions, and emotions are not just like ordinary states of the world in which we can happen to be or not, and in which we can sometimes bring ourselves, then maybe we should not say that they are *states* at all, but rather some other kind of thing that can be controlled in the way actions are controlled. Chrisman (2016)

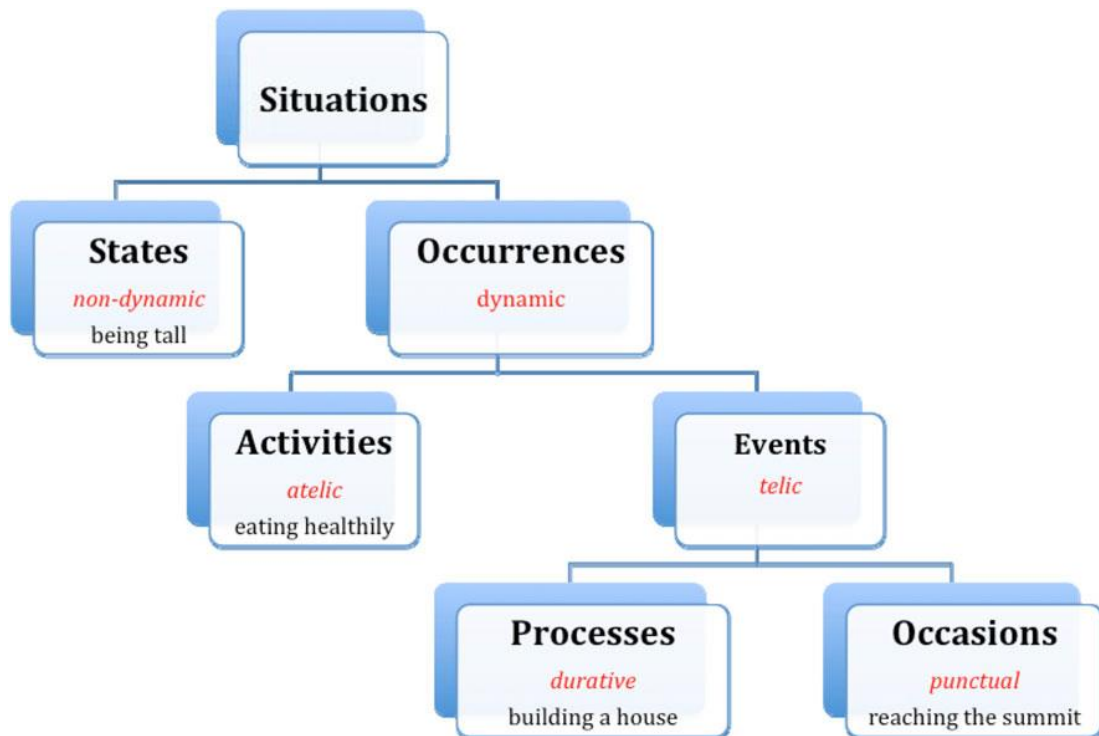


Figure 1: Vendler/Mourelatos typology

recently provided a good survey of accounts which approximate beliefs to more active kinds of things and highlighted some of their main problems. In what follows, I will first consider some *prima facie* plausible options and point out their main problems (9.1-9.2). I will then go on to discuss especially intentions and decisions, which seem to be most akin to action in the way we control them (in 9.3). It will turn out that, though the reasons-responsiveness of attitudes might provide *prima facie* grounds for thinking that attitudes are free or controlled in some *sui generis* sense, it does not provide support to the idea that we can voluntarily control our attitudes – that is, believe, desire, intend, or feel for practical reasons.

For the start, it is helpful to consider a typology of situations used by Chrisman (2016) (figure 1).¹ According to what I assumed throughout this book, attitudes fall into the category of *states* – like the state of being tall. It does not make much sense to say that somebody ought or ought not to be tall.² It might thus seem that if attitudes are just states, then it does not make much sense to say that one ought or ought not to believe, desire, feel, or intend something. A crucial difference between attitudes and these typical states, however, is that only the former

¹ The typology was developed in Vendler (1957) and Mourelatos (1978) and is derived from Aristotle's typology.

² It makes sense to say this, however, if we understand these sentences as state-'oughts' (see chapter 7.1) and assume that the person has some control over how tall she is: imagine a video-game in which you can adjust the size of your avatar depending on the situation – 'You should be taller when you come to this point of the game!' However, such state-'oughts' do not underwrite the idea that we are directly responsible for our attitudes in any way.

can be evaluated as rational or irrational. These evaluations define attitudes as a special subclass of states: those states which are subject to certain evaluations due to their presence in the space of reasons. Attitudes are those kinds of states which might pose initial problems for a naturalistic world view: states which have content, which are directed at the world, and which are evaluable in virtue of them having intentional content.³

The picture of attitudes as states seems quite coherent. What could be the reasons for or against putting attitudes in another ontological category? Take first the category of *activities*. If attitudes were activities, then the prospects of believing, desiring, intending, and feeling for practical reasons might be much better. For it seems that we can eat healthily for such reasons (see the example in the typology above). ‘Eating healthily’ is here to be understood not as a mere event, something we do at one point and then never again, but rather as an ongoing engagement with no specific end point (‘atelic’). If we conceive of attitudes as *dispositions*, it is not too far-fetched to conceive of them as activities. The activities that make up, for example, feeling angry, would then be the kinds of activities that are characteristically connected to this attitude – probably shouting, smashing things, or silently and grumpily staring at someone, or arguing with someone.

As I argued in chapter 7.2 when I discussed Rinard’s argument for practical reasons for belief, we do not perform long-term activities for reasons in the same way in which we perform basic actions for reasons or believe on the basis of evidence. If believing was a long-term activity, then our direct responses to reasons would be the smaller activities that make up believing. These things would then be what we do directly for practical reasons – not believing.

However, some might not be convinced by my point. For if believing can be conceived as a long-term activity, then we can believe for practical reasons in the same way as we can eat healthily, or live in the country, for practical reasons. Surely, one might argue, this sense of ‘doing something for a practical reason’ can ground a much more interesting form of pragmatism than the non-traditional pragmatism I endorse. The objector would grant to me that it is not interesting to show that we can believe ‘for practical reasons’ in the same sense as we can be in non-reasons-responsive states ‘for practical reasons.’ Yet, they would argue, the sense in which we can perform long-term activities ‘for practical reasons’ is a different sense that allows for formulating a different, and more interesting, version of pragmatism.

I am not sure why this version of pragmatism would be more interesting. I have argued in chapter 8.1 that we cannot perform long-term activities for reasons in the sense of ‘ ϕ ing for

³ See, for example, Papineau (2008).

a reason' in which believing for practical reasons would pose a problem for evidentialism. If beliefs and other attitudes were long-term activities, we would just have another, uninteresting, sense in which we can 'believe etc. for practical reasons.' Yet, for those who do not agree with me that this sense is uninteresting, I will argue in the following that it is implausible to conceive of attitudes as long-term activities. It follows that we cannot believe for practical reasons in the sense in which we can perform long-term activities for practical reasons.

One problem with an understanding of attitudes as activities is that we can conceive of someone having an attitude even if this attitude does not manifest itself in *active* behavior. Rather, our attitudes often manifest themselves involuntarily, for example, in trembling or blushing. Another worry is that some attitudes – such as the belief that there are stones on the moon, or the intention to leave our galaxy as soon as this is possible – need not *actually* manifest themselves in behavior at all in order for them to be ascribable to us. We just need to be *disposed* to manifest these attitudes in either active or passive behavior. Furthermore, we still retain many of our attitudes even when we are asleep, and thus do nothing at all – we do not stop believing what we believe while sleeping (Boyle 2011, 6).

A possible move would be to restrict the attitudes that are activities to those that are *actually* accompanied by actions we perform for reasons. We can form *these* attitudes for practical reasons, even though not the ones that are not accompanied by actions. Thus, we can believe for practical reasons that god exists insofar as we perform the actions that make up this attitudinal activity – praying, going to church – for such reasons. We can [intend to have a drink] for practical reasons by going to the bar, because our intention partly consists in going to the bar.

There is a serious problem with this account of attitudes as activities. Note first that attitudes are often mentioned in *explaining* why we act the way we act. The fact that a person believes in god (partly) explains why they are praying and why they are going to church. How could this fact do this explanatory work if the person's belief is *composed* of the actions it is supposed to explain? Chrisman also highlights this explanatory role (as well as the attitude's role in *predicting* what one will do), and its being conditional on viewing beliefs as *states* in which we are, rather than as things we perform:

[P]hilosophers standardly talk about *states* of belief not *performances* of believing, and this appears to be no mere façon de parler. [...] There's considerable utility in thinking of the mind as embodying various enduring and stable states, having information in the hard drive, so to speak, even when it is not actively being utilized. Combined with a conception of *desires* as states with a different and

interlocking functional role, we have a very powerful model for predicting what agents of various stripes will do and interpreting what agents are doing or have done. (2016, 7, emphases original)

It seems to me that we lose ourselves in a merely terminological debate if we understand attitudes as being *partly* active. If we *stipulate* that an attitude is a thing that involves actions as constitutive parts of it (let us call such things attitudes*), then I have no trouble allowing that we can perform these attitude*-constituting actions for practical reasons. It seems apt to say that we can be angry* for practical reasons given this concept of anger*: we can perform the actions that are part of being angry for such reasons. However, and this is the crucial point, this stipulation would not transform the *involuntary* parts of anger* into ones which we can have for practical reasons. If, by ‘anger*’ we understand also the actions that are involved in being angry*, then *this part* of anger that consists in performing actions can be done for practical reasons.

I do not think that this use of ‘anger*’ is our ordinary use of ‘anger.’ More importantly in this context (but probably resulting from my point about ordinary language), it is not the use of ‘anger’ we wish to analyze when we ask ourselves whether we can have attitudes for practical reasons. When we ask this, we are motivated by the thought that our *involuntary nature* is subject to norms – that we can be blamed and praised for responses we do not control in the way we control actions. If we use ‘anger’ in this new way as involving actions, then we must reframe the problem: what makes us responsible *for the part of anger* that is not an action*? Here, the problem that we cannot control this part in the way we control our actions just arises from anew.

Thus, given that we cannot avoid a passive element in our analysis of the nature of attitudes – due to the explanatory role they need to fulfill, and due to our interest we have when we ask whether we are responsible for our attitudes –, we cannot preserve the idea that we can have attitudes for practical reasons by making them partly actions. For saying that we can perform these actions for practical reasons does not help us in answering what is ultimately of interest: can we have the *passive* elements of the attitude for practical reasons?

The same argument applies, *mutatis mutandis*, to attitudes as dynamic processes (building a house) and as dynamic events (reaching a summit). Given the explanatory role we need attitudes to fulfill, and given our interest in discussing them here, we cannot refer to them as something that is completely active. Introducing active elements as components of attitudes will not imply that the passive elements are something we do or have for practical reasons.

When building a house, part of it might be that we are sweating. But the fact that building a house is something we can do for practical reasons does not imply that our sweating, which is part of it, is something we can do for practical reasons.⁴ Similarly, when we ‘believe in god’ (understood as a dynamic process), part of it might be that we go to church. But this does not explain why the *state* involved in believing, which explains our action, and in which we are interested in the present inquiry, is something we can do (or be in) for practical reasons.

Similarly, when we reach a summit, we might have done so for practical reasons (say, we wanted to prove ourselves that we can still do it), but this does not imply that some passive movement in our body that is going on while we perform the actions involved in this event of reaching a summit is something that we do for practical reasons.

One might object that ‘forming an attitude’ can – understood in a certain way – belong to the same category as ‘reaching the summit.’ In this sense, ‘forming an attitude’ would be a punctual and non-durative event, and yet we might conceive of it as an action like ‘reaching the summit’: something we can do for practical reasons.

My answer to this objection is twofold. First, I have already argued in chapter 8.3 that to influence our beliefs indirectly by any ‘formings’ that are different from the actual belief does not establish voluntarism, and thus not pragmatism. If these formings were punctual and non-durative actions, then we could at most do these *formings* for practical reasons, but not the resulting belief. Secondly, however, I do not think that ‘reaching the summit,’ understood as a punctual and non-durative event, is an action we do for practical reasons. Rather, it is the result of our efforts. Insofar as ‘reaching a summit’ is an action, it is not to be understood as a punctual, non-durative event.⁵

I conclude that the fact that active elements of attitudes – if there are any – can be done for practical reasons does not imply that the passive elements, which are at interest here, can be done for practical reasons.

⁴ One might doubt that ‘sweating’ is part of building a house. However, insofar as we want to understand attitudes – which, as I have argued, include a *passive* element – as a dynamic process like building a house, we need to allow for passive elements in the attitudinal dynamic processes. That we can perform the active elements in the processes for practical reasons does not imply that the passive one’s can be done or held for practical reasons. This general point is not affected by my concrete example.

⁵ See my points in chapter 8.2 on the concept of finding. We might want to refer by ‘the scientist is finding a cure for cancer’ to her *action* of moving around in excitement before she actually finds the cure. But it is clear that in this use of the word, it is an ordinary action. If we use ‘forming a belief’ as referring to an action prior to the actual formation (e.g., the action of reading a newspaper), then this formation is trivially an action. But the fact that this action is active does not make the resulting belief something that is done for practical reasons.

9.2 ACTIVE ATTITUDES AND INDIRECT CONTROL

Let us now, in light of this argumentation, have a look at a concrete proposal for understanding attitudes as active (i.e., as on the right side of the typology in figure 1). According to Boyle, beliefs consist in ‘actively being a certain way’ (2011, 19), that is, in ‘an exercise of agency that does not take the form of an occurrent process or event’ (ibid., 16). Note first that it is hard to locate this suggestion in our typology above (figure 1). It seems that Boyle argues for introducing a category of thing that neither fits with the right side of dynamic performances, nor with the left side of static states. As Chrisman (2016, 9-10) points out, Boyle introduces an ‘active state.’⁶

What exactly could Boyle mean? In line with the argument just stated, we can distinguish between two claims his theory could amount to:

- (a) Beliefs are partly active, that is, they involve activities, but they also involve passive states or dispositions.
- (b) The constituents of belief that are traditionally conceived of as passive, that is, the state- or disposition-element, is in fact in some sense active.

As I have argued above, if Boyle’s claim amounts to nothing more than (a), then it does not answer the question of how the passive aspects of belief can be done for practical reasons, or, for that matter, how it can be something for which we are responsible. Thus, the interesting reading is (b). On this reading, however, I have a hard time making sense of it.

It is important to note, first, that Boyle does not aim at establishing that we can believe for the same reasons for which we can act – beliefs are not meant to active in exactly the same sense as actions are active. It would indeed be hard to understand how states can be active in the same sense as actions are. What Boyle wants to explain is why beliefs are subject to *epistemic* norms, rather than practical norms. According to him, this requires a form of agency. This assumption, however, rests on the false idea that epistemic standards are requirements analogous to requirements of action (see part II, esp. the ‘Summary and Conclusion of Part II’). If we are not blameworthy for failing to comply with epistemic standards in any sense of ‘blameworthy’ that presupposes control, then the assumption that these standards presuppose agency is unwarranted.

However, most importantly for our purposes here, the agency Boyle introduces is *sui generis* – it is not the ordinary agency that we also exercise when we perform full-blooded

⁶ Hieronymi (ms) also defends attitudes as belonging to a category of active states.

actions for reasons. Rather it is a kind of agency we are supposed to exercise when we believe for *epistemic* reasons. His account thus will not, and is not meant to, support the idea that we can believe for *practical* reasons.

Since I here mentioned the idea of direct control over attitudes or attitudes as somehow being ‘active’ (an idea of which I am suspicious), I will use the opportunity to make some brief remarks on mental agency throughout the end of this subchapter and the next one in order to frame how I conceive of the way we exercise agency over our attitudes, before I return to my case against mental voluntarism. These remarks are not essential to my argument against believing at will. Yet I think it might be interesting to sketch my overall picture of mental agency already at this point of our inquiry, thereby shedding some light on the face of mental responsibility that presupposes indirect control (according to the account I will return to in part IV). Chapter 12 will elaborate on this understanding of indirect control I will now sketch.

In chapter 8.3, when considering objections to my argument against the possibility of forming attitudes at will, I already mentioned an objection to accounts that say that we are active *in believing* itself. I pointed out that when we start to think more thoroughly about the nature of this activity, we tend to construe *further* activities that are different from believing, and by means of which we come to believe, or maintain our believing. We can see this in Boyle’s account, too. This is obvious already in the way he formulates his central question: ‘[W]hat does it mean to say that we can ‘make up’ our minds? In what sense are judging and choosing things that we do rather than things that merely happen to us?’ (Boyle 2011, 2). Thus, Boyle asks himself whether we are free in *judging* that something is true – but judging is not believing (*ibid.*, 5). Rather, judging is usually conceived of as the production of a belief, or, more in Boyle’s vein, as an activity that somehow goes hand in hand with believing. Neither the first option nor the latter helps us with locating agency *in believing itself*.

One might wonder about the picture I defend when I suggest that our control over attitudes is only indirect. Boyle, Chrisman, Hieronymi, and many others agree that if we do not locate agency somehow in *believing itself*, or at least connect agency closer to belief than my picture of indirect doxastic control suggests, agency over belief is *too external* to belief to make sense of mental responsibility. Boyle claims that if an account locates agency merely in acts preceding the belief, like deliberation and judgment, then ‘our agency can get no nearer to our beliefs than to touch them at their edges’ (2011, 6). The question is, of course, whether this is something that should worry us. Part of the worry derives from the idea that our beliefs are subject to the standards of epistemic rationality, which are norms of *belief*, rather than of something else that is closely connected to belief. The thought that these norms presuppose agency,

however, only comes up if we think that we are responsible for whether we comply to them in a way that presupposes control. I have already pointed out at various points in this book, and will argue at length in part IV, that this idea is misleading.

However, there is still another worry I did not fully address until now that might motivate the idea that we must exercise agency in believing. It seems that we are responsible for our *own* beliefs in a special way – a kind of responsibility that is more direct than the one we have for the attitudes of others, or for ordinary states of the world we can influence through our actions. In response to this worry, I will, first, argue in chapter 12 that this idea can be accounted for, even if we accept that believing is not itself an exercise of agency, and that the only way we can control our beliefs is by means of ordinary actions that are not themselves part of the belief. Furthermore, I argue in chapters 10 and 11 that there is a sense in which we are directly responsible for our beliefs and other attitudes that does not presuppose control at all.

Chrisman objects to a view that locates agency merely in activities previous to a belief that ‘normal deliberation doesn’t have predictable effects such as believing *that it’s raining*; usually the best we can predict is that we will form some doxastic attitude or other about the question under deliberation,’ and concludes from this that ‘if all of the agency in the vicinity of belief is in deliberation and judgment, then it looks as if we exercise cognitive agency with respect to belief only very indirectly and rarely’ (2016, 5). Again, the question is that, even if we accept his conclusion, whether this should be troubling. If our aim is to explain why we are responsible to *epistemic* norms, then this indirect kind of agency will not do (see chapter 4.4). With this, Chrisman is right. Epistemic standards govern our beliefs more often than we are able to exercise indirect control over them: they also govern beliefs that are held as results of subconscious reasoning mechanisms, for example. Furthermore, they are standards of *believing*, and not of *acting*. However, again I do not think that there is a need to assume any kind of agency in order to explain our responsibility to such standards. If these standards do not *require* beliefs in the same way as requirements of morality *require* actions (see esp. ‘Summary and Conclusion of Part II’), then they might be in place even if it is not under our control whether we comply with them.⁷

⁷ The premises of Chrisman’s argument are also doubtful. Meylan (2017) argues plausibly that indirect doxastic control is much more often exercised as it is ordinarily supposed. I agree that if we want an understanding of epistemic reason as placing *requirements* on us that are similar to the requirements morality places on us, then the only way we can get this is by viewing epistemic norms as norms for activities associated with belief (she spells out her account also in Meylan 2013). However, such ‘epistemic norms’ would not be properly epistemic in my sense. Rather, they are practical requirements of belief-management. In the end, Chrisman (2016) provides a similar picture of epistemic norms as Meylan does, only that he adds the activity of ‘maintaining a coherent system of beliefs’ *while* believing (to the activities by means of which we can influence beliefs previously to

9.3 FREEDOM OF INTENTION AND DECISION

Note that the argument I spelled out throughout 9.2 also applies to intending and deciding. Though many authors argue that intending or deciding is a paradigm of an exercise of agency, they also think that the agency involved in intention and decision must be different from the agency involved in acting.⁸ Insofar as having an intention involves having dispositions to act a certain way under certain circumstances, and thus can be employed in explaining why one acts the way one does, our ordinary concept of intention is a concept of a *state* of mind.

How does decision fit into that picture? We come to intend to do something by *deciding* what to do, as we come to believe something by judging that *p* or forming the belief that *p*. We decide to do something either by deliberating about what to do and then settling the question what to do, or by just setting out to do it. The latter way of deciding explains why it *seems* that we have direct control over our decisions: we can decide to do something *by* setting out to do it. For example, by getting out of bed in the morning I can make it the case that I have decided to get out of bed. I here exercised direct control over my action of getting up rather than over my decision of getting up. The decision is only something I had under my control in virtue of my performing the action for which I have decided. In this vein, Soteriou (fc) argues that we can sometimes make it true that we exercised our agency in deciding if we later actualize our decision in action: it seems as if our control over decisions is somehow conceptually dependent on the control we exercise over the actions for which we have decided, and that are temporarily *later* than our decision.⁹

Yet a decision is not itself an action. We cannot decide something just because it would be pleasurable to *decide* it (rather than because it would be pleasurable to do what one decides). And yet a decision cannot be conceived of as a *state*, like an intention. A decision happens in

their existence). Both accounts, however, reduce epistemic norms to norms of action, and thus I do not think that they capture what most epistemologists usually mean when they talk about justified or unjustified, rational or irrational beliefs, which are evaluations that are supposed to apply *directly* to belief. Epistemic justification or rationality, in the common use in epistemology, is only the fittingness of belief to the evidence, and thus no requirement analogous to the requirements of morality, as already argued throughout part II of this book.

⁸ Cf. the accounts of Hieronymi (2014), McHugh (2013a), Owens (2000), and Shah (2008). For the classical case that intention is not as free as an action, cf. Kavka 1973.

⁹ It is important to note that I do not claim that we can *cause* our past decisions by future actions: I do not commit myself to any backward-causation (and neither does Soteriou). The idea that we ‘can make it true’ that I have decided to perform an action by performing that action becomes plausible as soon as we realize that decisions are not objects in our mind, but rather more like theoretical postulates we assume in order to explain behavior. This does of course nothing to undermine the reality of decisions: for in order for them to explain an action, they must exist.

an instant, but states are persisting over time. A decision is an occurrence we can bring about, similar to the way a formation of an attitude is an occurrence we can bring about (or how ‘finding something’ is something we can bring about). It seems that a decision *just is* the formation of an intention – or, more precisely, it is a formation of an intention in the passive sense of ‘formation’ where a formation is *not* identical to deliberation, but rather consists in the change of events from ‘I did not decide’ to ‘I decided’ (see chapter 8.2).

Intention and decision thus do not deserve any special treatment when it comes to our agency we exercise over them as compared to the agency we can exercise over other attitudes or attitude-formations. Rather, an intention is as active or passive as a belief. A decision is as active or passive as the formation of a belief, or a judgment that something is the case. Let us now return to my case against mental voluntarism, and thus against mental pragmatism.

9.4 GUIDANCE CONTROL OVER ACTIONS AND ATTITUDES

McCormick’s version of voluntarism also locates agency *in believing itself*, rather than in the formation of a belief, or in some kind of activity external to believing itself (McCormick 2015, chapter 6; 2017b). McCormick might thus escape my argument from chapter 8.2 by rejecting the assumption that believing at will means to *form* a belief for practical reasons. Rather, it just means *to believe* for practical reasons. In this section, I first spell out the concept of *guidance control* that is central to her account and show how McCormick argues for the idea that guidance control underwrites the idea of doxastic responsibility to practical norms, that is, for doxastic pragmatism. The first trouble I point out for her position is that mental states cannot be related to reasons-responsive mechanisms in the same way as actions can. The second trouble is that the guidance control approach to responsibility (for actions or mental states) leaves the connection of guidance control to voluntary control in the dark. Both troubles will call into question whether guidance control is suitable for underwriting the subjection of mental states to norms of action: it will call into doubt whether McCormick’s strategy for establishing pragmatism is promising.

In the context of my overall dialectic of this chapter, I read McCormick as proposing that we can believe for practical reasons *although belief is ontologically different from action*. My claim is not that this is her explicit position. If it is not, then she will be confronted with the troubles I presented in subchapters 9.1. and 9.2, where I argued that at least the passive element necessarily involved in believing (or desiring, feeling, intending) is not itself an action. My discussion of McCormick’s argument will highlight two general problems with positions

that allow attitudes to be ontologically different from actions, but at the same time allow for practical reasons for attitudes. Before presenting the problems, I introduce her account (9.4.1). The first problem is that it is hard to see how two kinds of things which are ontologically very different can be controlled in the same way (9.4.2). The second problem is a problem for an attractive strategy of making sense of the idea that actions and attitudinal states are controlled in the same way (9.4.3). The strategy is to introduce a second kind of control that we have both over actions and attitudes (e.g., McCormick's 'guidance control'). However, introducing a second kind of control leaves it mysterious how this kind of control is related to the agency we also exercise when we act for practical reasons ('voluntary control').

9.4.1 GUIDANCE CONTROL AND PRACTICAL NORMATIVITY

According to McCormick, beliefs as well as actions are sometimes under *guidance control* – a concept introduced by John M. Fischer and Mark Ravizza (1998) into the debate about responsibility and determinism. Fischer and Ravizza already claimed that their basic idea is not restricted to responsibility for actions and their consequences, but can be applied to responsibility for emotions, since we also have guidance control over emotions (Fischer/Ravizza 1998, Appendix). This suggests that the notion of 'guidance control' might not only explain our responsibility for actions or beliefs, but also for (some or all) other mental states as well. It seems to be a suitable candidate for this task, because, as we will see, the conditions something must satisfy in order to be subject to guidance control are unproblematically fulfilled by beliefs (McCormick 2011, 174-176; 2015, ch. 6), and also by other mental states. According to McCormick, the fact that we have the *same kind of control* over both action and belief is meant to support voluntarism, that is, the idea that we can believe for the same kind of reasons for which we can also act (i.e., practical reasons).

We have guidance control over an action or mental state if and only if 'the mechanism that actually issues in the action [or mental state] [is] one's own, and the mechanism [is] responsive to reasons' (Fischer/Ravizza 1998, 89, my add). Examples for the relevant reasons-responsive mechanisms, that is, for the reasons-responsive 'process[es] that lead[...] to the action, or the "way the action comes about" ' (Fischer/Ravizza 1998, 38), are, with respect to action, practical deliberation and habit, or, with respect to belief, perception, memory, and deliberation about what to believe (McCormick 2015, 113-117).¹⁰ Since both action and belief

¹⁰ Furthermore, inquiry, evidence gathering and attention are mentioned by McCormick, but not discussed in detail. I will not go into the details of what reasons-responsiveness is. The main idea is that being reasons-

can result from reasons-responsive mechanisms, McCormick claims that they are controlled in the same way. By viewing our theoretical freedom in analogy to our practical freedom in this way, McCormick claims to provide a unified picture of what were traditionally supposed to be *different* forms of normativity.

The ownership-condition on guidance control is supposed to avoid the absurdity that we can be responsible for mechanisms which were too much influenced by factors beyond our control. It is, for example, conceivable that an evil scientist of the future is implanting a reasons-responsive mechanism in someone. If the first condition was sufficient for being responsible for X, then the person would be responsible for the results of such mechanisms as soon as they were implanted into them. This is an absurd consequence, according to Fischer and Ravizza as well as McCormick, and thus we need the condition that the person has to *own* the mechanism. More mundane examples will also do. There may be cases where, for example, our reasoning is subject to external influences to such a degree that it undermines our responsibility for the resulting belief. The fallacious reasoning of a racist ‘This person is a member of group X, thus she will probably do something illegal soon,’ may not lead to a *blameworthy* belief if the belief was influenced by a racist community in a certain way: given that the racist is very young and did not yet have enough ‘practice’ in reasoning while at the same time being subject to social pressure, we may (in a sense) excuse her implicit belief ‘that members of group X are doing illegal things all the time.’ Even though the subject has a morally objectionable belief, we may explain our appropriate withdrawal of blaming responses like resentment or indignation by saying that the subject did not really *own* the mechanism from which the belief resulted, and thus we cannot say that she is responsible for having it.¹¹

responsive or having a reasons-responsive mechanism is necessary to be free in the sense required for full responsibility. As an example of someone who is intuitively not reasons-responsive, McKenna (2016, 27) mentions a compulsive hand-washer, who washes his hands in a particular scenario because they are dirty. Due to his compulsion, he would wash his hands in most close counter-factual scenarios where they are not dirty, and thus his action, even if done for a reason, is not the result of a reasons-responsive mechanism. It might be doubted, however, that the hand-washer can be said to wash his hands for the reason that they are dirty if he would not do so in close counter-factual scenarios where they are not dirty.

¹¹ Smith argues that in such cases, the subject is still responsible for her belief insofar as it reflects an evaluative judgment about her available evidence, and explains the intuition that we find the belief excusable by saying that the subject is not responsible for becoming the kind of person she is (Smith 2005, 267-268). We might want to say that we are, in a sense, *always* responsible for our beliefs insofar as they are full-blooded attitudes within the space of reasons. I return to such issues in chs. 10 and 11, and I will argue that Smith is right insofar as she is talking about a specific *kind* of responsibility that is not the proper basis for such passionate blaming responses like resentment or indignation. However, it is important to note that the mere fact that beliefs are *bad* (whether morally, rationally, or otherwise) does not allow for the conclusion that we are always responsible for them: mere evaluation is not sufficient for responsibility. In some sense of the word ‘blame’ – namely, when we think about it in terms of emotions like resentment, indignation, or guilt –, blame becomes inappropriate as soon as external influences, in contrast to the use of someone’s own reason, are the main explanation for why someone

In order to *own* a mechanism, we need to show in the course of our life that we take responsibility for it (McCormick 2015, 112-3). Taking responsibility means to view yourself as an active being able to influence your own environment, and as the proper object of reactive attitudes. Such a perspective on yourself manifests itself in certain ways of behavior (in a very broad sense of ‘behaviour’ that includes emotional reactions) which show that you have taken responsibility for the relevant mechanism. For example, in the case of belief, you may feel regret, or even guilt, after someone pointed out to you that you misremembered, say, your friend’s birthday. Behaving in such a way makes you responsible for the doxastic output of your memory mechanism and shows that you are aware of your own responsibility for the resulting beliefs.¹²

The emerging picture is the following: We act and behave in a way which reveals our responsibility for certain reasons-responsive mechanisms, which produce actions as well as mental states. Even though we are not influencing the products of our mechanism by (other) actions at the moment of their production, we can nevertheless be said to have guidance control over the product of the mechanism, and are thus responsible for the result.

How does McCormick use this approach to argue for doxastic pragmatism? I think her main argument can be stated as follows:

- (1) Actions are subject to practical norms because we have guidance control over them.
- (2) If actions are subject to practical norms because we have guidance control over them, then anything over which we have guidance control will also be subject to practical norms.
- (3) We have guidance control over beliefs.
- (4) Thus (from (1) to (3)): Beliefs are subject to practical norms.¹³

has a belief. In chs. 10 and 11, I will argue that there is *a sense* in which we are always responsible for our attitudes, even if they were wholly beyond our control (cf. also Adams 1985).

¹² McCormick emphasizes that we can be wrong about whether we own a mechanism: ‘If one were being directly manipulated – for example by electronic stimulation – one’s feeling like one was an appropriate target for reactive attitudes would not thereby make one responsible’ (McCormick 2015, 121). We cannot just escape our responsibility by not viewing ourselves as responsible. In addition, we pay a high price for not viewing ourselves as responsible: We exclude ourselves from the moral community (at least in our head).

¹³ I take this line of thought from McCormick’s emphasis on the parallels between the ways we control (some) actions and (some) beliefs. She clearly seems to be committed to (2), given her idea that guidance control underwrites our direct responsibility we have for both: ‘[Y]our responsibility for your beliefs is as direct as responsibility for many of your actions; in each case you are responsible if the action or belief resulted from a mechanism for which you have taken responsibility’ (McCormick 2015, 122). This kind of direct responsibility is, in turn, seen by her as a precondition for subjection to practical norms. It is also clear that she accepts (3) and (4). For reasons of logical consistency, I thus interpret her also as accepting (1).

I read ‘being subject to practical norms’ here as ‘being an entity which we can do or have for practical reasons.’ I take (3) for granted. One can object to (2) that the way we control states by guidance control is *different* from the way we control actions by guidance control, and because of this, guidance control cannot underwrite the subjection of mental states (including beliefs) to practical norms. Furthermore, one can object to (1) that guidance control does not even underwrite the subjection of *actions* to practical norms. So why should it underwrite the subjection of other things to practical norms? I will spell out these two worries in the following subchapters. Both will lead us to more general problems with the idea that attitudes could still be subject to practical norms even if we conceive of them as ontologically different from actions.

9.4.2 ACTIONS AND MENTAL STATES AS RESULTS OF MECHANISMS

What does it mean to say that something ‘results from a reasons-responsive mechanism’? One way to interpret this phrase is by taking ‘results’ as referring to *consequences* in a broad sense. For example, if I reason about whether it will rain, one consequence of my reasoning might be that I believe that it will rain, while another one might be that I have a headache (say, I reasoned very hard). In this first sense of ‘result,’ both my belief and my headache are results of my reasoning.

It is obvious, however, that my belief results from my reasoning in a different *way* (or *sense*) than my headache. While the content of the belief can be seen as the conclusion of the contents of the attitudes I reasoned with, my headache cannot. I could have reasoned: The heaven looks like it will rain, so it will rain. But I cannot have reasoned: The heaven looks like it will rain, so I have a headache. The second ‘line of thought’ cannot count as reasoning because the fact that the heaven looks like it will rain is no evidence for me having a headache, but for rain. Thus, there is a second, and more restricted, sense in which mental states can ‘result’ from mechanisms. In this sense, we can say that beliefs result *directly* from my reasoning, whereas mere consequences or side-effects of the operations of my mechanisms will not count as direct results.¹⁴

¹⁴ This does not only hold for *reasoning*-mechanisms. Also my *perception* may lead me, first, to a belief about my environment, and, secondly, to have a headache (say, I looked into bright light and I am quite sensitive to light). Obviously, the sense in which my belief ‘results’ from my perception is more restricted to the sense in which my headache results from my perception. My belief is no mere consequence or side-effect of my perception.

Since McCormick argues that we control actions and beliefs in the same way, we would expect actions and mental states to *result* from reasons-responsive mechanisms in the same way (or sense of ‘result’). But they do not. First, my action will not be a mere consequence or side-effect of my reasoning like my headache above, but rather an actualization of the conclusion of my practical reasoning (that is, an exercise of my intention). Even if we would allow that my action was a mere consequence of my reasoning, this would not help McCormick. If *this* was the sense in which actions and mental states result from mechanisms in the same way, then we would treat actions as on a par with mere consequences of actions. We are not directly responsible for mere consequences (as we are for most of our actions), because we do not control them directly (as we control most of our actions).¹⁵ Such indirect control is not sufficient for believing for practical reasons. This is because controlling a belief only *indirectly* by means of a previous action will never amount to believing for practical reasons,¹⁶ and thus indirect control over belief would not help with establishing McCormick’s version of doxastic pragmatism. If you think that it is better to believe that p, and you perform certain activities which bring you into the state of believing p, you do not *believe* for practical reasons, but only *act* for practical reasons.

Maybe actions can result from mechanisms in the same way as mental states do if we interpret ‘result’ in the second sense identified above? Actions, however, do not result *directly* from reasoning, like beliefs or other mental states to which we can reason (including desires, intentions, and, I think, many emotions). This is shown by the fact that when we reason about what to do, our reasoning, if it comes to a successful end, will not conclude in an action (whatever that could mean), but in an intention, which in turn must be executed in order to issue in an action:

You can come to act in a certain way by reasoning to the intention to act in that way. [...] For all to go well you must retain the intention, and successfully execute it. Execution per se is not a job for reasoning, but for the will, and for example, certain motor systems (McHugh 2014b, 14).

As Andrea Kruse points out, it is confusing to claim that the guidance control approach of Fischer and Ravizza explains our freedom and responsibility merely in terms of mechanisms. ‘A process or a mechanism can neither *recognize*, *choose* nor *do* anything. These attributes

¹⁵ I say ‘most of our actions’ because I granted in chapter 7.2 that we do not respond as directly to our practical reasons when we perform long-term activities as when we perform basic actions.

¹⁶ Cf. my argument against Rinard and McCormick throughout chapter 7. For some other proponents of this claim cf. esp. Kelly 2002, 174-176; Owens, 2000, 29-35.

apply to the agent level and not to the level of processes' (2016, 10). As I will argue in subchapter 9.4.3 below, leaving out the agent of the picture in this way seems to leave out an essential aspect of our freedom: our ability to exercise voluntary control. For now, it is only important to see that actions are not just direct results of mechanisms, like our attitudes, but rather require *more* – usually some kind of effort of will by the agent – in order to 'result' from a mechanism.

One could object that in cases where my action is *not* preceded by an intention, my action-producing mechanisms can result directly in an action. After all, McCormick does not think that we control beliefs in the way in which we control actions to which we have *explicitly deliberated* by thinking about practical reasons. Rather, she refers on more spontaneous actions, like the steps we take when taking a walk (McCormick 2015, 82). As Alfred Mele argued, such actions do not seem to be preceded by an intention (Mele 1997, 242-3).

There are two ways open to reply to this objection. First, I could deny Mele's claim and say that intentional activity is always execution of (previously held) intentions, however implicit or unconscious those intentions might be. However, secondly, I do not have to commit myself here to the claim that every action is preceded by an intention, and so I could accept Mele's point without contradiction. Even if there are mechanisms which cause actions without previously issuing in an intention, these actions still do not result from the mechanisms in the way an intention would. First, intentions (as well as beliefs) are still *states* and not themselves actions. Why should a state result from a mechanism in the same way as an action? McCormick does not argue that states are themselves actions (and I argued that they are not in chapter 9.1). But as long as both kinds of things belong to very different ontological categories, it is not clear that both will result from mechanisms in the same way. There are two independent reasons for this claim.

First, the agent needs to come into the picture and do something in order to act for a reason, but the agent does not need to *do* anything when an attitude results from a reasons-responsive mechanism. Furthermore, intentions have content (you intend *to do X*), while actions do not. Intentions can thus be conclusions of reasoning and, in this sense, *directly* result from reasoning, while spontaneous actions cannot insofar as they cannot be the conclusions of reasoning in the same sense. There are of course other mechanisms than reasoning from which spontaneous actions may result. However, it is not clear why actions should be capable of resulting directly from another mechanism than reasoning if they cannot result directly from reasoning.

Let me summarize my argument. I have granted throughout my discussion that guidance control *is* a genuine form of control (the next subchapter 9.4.3 will throw some doubts on this). Since whether we have guidance control over something depends on the nature of the mechanism from which it results, it is a plausible assumption that a difference in the way beliefs and actions result from mechanisms indicates a difference in the way we control both via exercising guidance control. Given this assumption, it follows from the differences I pointed out that guidance control over action cannot be the same as guidance control over belief (or other mental states, for that matter). By applying the idea of guidance control to belief, McCormick just shifts the problem to another level: while the initial trouble was how to make sense of the idea that we control both actions and beliefs in a similar way, we now have trouble making sense of the idea that both kinds of things result from mechanisms in the same way.

Let me stress my point again in other words. It is true that both actions and beliefs result from mechanisms which are responsive to reasons (and which are our own in the relevant sense). If this is sufficient for them to be under guidance control, then we *have* guidance control over actions and beliefs. But it does not follow that we control both kinds of things *in the same way*. Just *calling* both kinds of control ‘guidance control’ does not *make* them the same. For it seems that mental states result very differently from mechanisms than actions do. As long as both kinds of things belong to different ontological categories, it is not clear that both will result from mechanisms in the same way. Performing an action is a dynamic event, while forming or having a mental state is something static, or at least something to which I do not contribute to in the same way as I do when I perform an action. When thinking about how both kinds of things ‘result’ from mechanisms, we naturally tend to conceive of attitudes as somehow ‘popping up’ after the operation of the mechanisms. But actions do not just ‘pop up.’ Rather, we actively contribute to their upcoming and continuing performance. Or to put it another way: while we can *perform* actions, we cannot perform beliefs; while we can *bring about* a belief, we cannot bring about an action. Guidance control over mental states is not the same as guidance control over actions.

Thus, by drawing a parallel between the way we control beliefs and the way we control actions, McCormick does not establish that beliefs are subject to the same norms as actions. The concept of guidance control can be applied to mental states as well as to actions. But this does not mean that we control mental states in the same way as actions, because the way actions result from mechanisms is quite different from the way in which mental states result from them. This makes premise (2) of McCormick’s argument highly doubtful.

We can draw a general lesson from this discussion. I have already argued that attitudes are ontologically different from actions (9.1). Thus, I read McCormick as arguing that they can nevertheless be controlled in the same way – they can both be direct responses to practical reasons. This claim is confronted with the problem that when we think about how both kinds of things flow from our reasons-responsive mechanisms, we see that there is a huge difference in the way they do. This means – and this is my general lesson – that the way we can control both kinds of things must be very different. On the plausible assumption that believing for practical reasons requires that one controls beliefs in the same way as one controls other responses to practical reasons (that is, actions), it follows that one cannot believe for practical reasons.

9.4.3 GUIDANCE CONTROL, VOLUNTARY CONTROL, AND DIRECT RESPONSIBILITY

The fact that Fischer and Ravizza's approach can be so easily applied to beliefs (and emotions) might fire backwards. Rather than showing that beliefs (and emotions) are subject to practical norms, the applicability of guidance control to such states may make us doubt whether the approach works for actions in the first place. Maybe it gives us too passive of a picture of our active relation to our own movements – maybe it treats our actions more passive and state-like as they in fact are. Maybe guidance control is, after all, no form of genuine control. At the very least, it should make us doubt whether guidance control is the *whole* of control we have over our actions.

If *voluntary control* is a different form of control than guidance control, we will have some reason to doubt that guidance control can underwrite action's subjection to practical norms (i.e., McCormick's premise (1)).¹⁷ The fact that we have *two kinds* of control over (most of) our actions calls into question whether *one* of those kinds is sufficient to explain why they are subject to practical norms. The problem for McCormick's pragmatism I spell out in this section can be summarized by the following argument:

- (1) Sometimes we have guidance control as well as voluntary control over an action that is subject to practical norms.

¹⁷ Verena Wagner sees a difference between voluntary control and guidance control, too: 'What makes this account of guidance control to be a promising model for doxastic control is the fact that the agent is not required to exercise any *voluntary* control over her mechanism and the way she responds to reasons. The responsiveness of the mechanism that leads to action is not a capacity that the agent exercises through her will; [reasons-responsiveness] is better to be seen as a special *feature* of the process that results in the agent's willing and performing an action' (2015, 7).

- (2) Voluntary control is not the same as, nor reducible to, guidance control.
- (3) If we have two fundamentally different kinds of control over an action that is subject to practical norms, then it is not clear whether one kind of control is by itself sufficient to explain why the action is subject to those norms.
- (4) Thus (from (1), (2), and (3)): It is not clear whether guidance control is sufficient to explain why actions are subject to practical norms.

I will only comment in more detail on premise (2), because I think it needs most support. Premise (1) is not up to dispute as long as we have not yet decided whether guidance control and voluntary control are different, as premise (2) states. If they are just the same kind of control, then premise (1) simply states that we are active when acting. I do not comment premise (3), because it is formulated very carefully with the phrase ‘it is not clear.’ Whether (3) is true in stronger formulations that avoids this epistemic phrasing cannot be decided within reasonable scope, and is not necessary for making my point. I am satisfied here if I can show that there *might* be a problem for McCormick’s version of doxastic pragmatism (which relies on voluntarism) that my account of reasons for attitudes – constitutivism – can avoid. If we accept conclusion (4), then McCormick’s line of reasoning presented in the last subchapter cannot obviously go through, since (4) is the (epistemically careful) denial of her first premise (see subchapter 9.4.1 above).

To support premise (2), we need an argument why voluntary control is not reducible to guidance control. The ownership component in the notion of guidance control entails that voluntary control cannot be *the same* as guidance control. According to this component, being responsible for a mechanism implies that I exhibited behavior in the past that reveals my responsibility. Now, we cannot conceive of a scenario in which someone takes responsibility for a mechanism without performing free actions before taking responsibility. If we could not act freely before we first took responsibility, we would just be victims of the circumstances until then. How could such victims ever become responsible for anything? When children become members of the moral community and fitting targets of the different reactive and moral attitudes, they already had the ability to act freely. They were no mere biological automats, but rather active, even though not yet responsible, beings. Moreover, they become responsible *because* they can act freely. Fischer and Ravizza themselves admit this when they emphasize that integrating children into our moral community involves acting *as if* they were already responsible for what they do: when children unpack the presents of other children, although they are not allowed to do so, we might call their attention to the fact ‘that the opening of the presents

was a result of [their] own exercise of agency – the presents didn't just open “on their own,” or as a result of a freak natural accident, or mistake, and so forth’ (Fischer/Ravizza 1998, 208). Even if our attitude towards children involves acting *as if* they were responsible, though they are in fact not yet responsible, it does not involve acting *as if* they are free agents – they already *are* free agents in that they exercise voluntary control over their conduct: they can already act for reasons or execute their intentions.¹⁸

However, even if we accept that guidance control cannot be the same as voluntary control due to the ownership-component in guidance control, we may think that voluntary control can be *reducible to* guidance control. Such a picture could look as follows: Children are free in the sense of having reasons-responsive mechanisms from which their actions result without *owning* them in the sense necessary for guidance control. This avoids bringing a substantive notion of voluntary control into the picture of human agency: our freedom would be reasons-responsiveness, and full freedom and responsibility would require ownership in addition. Voluntary control would be just an element of guidance control, namely, responsiveness to practical reasons.

However, it is doubtful that voluntary control can be identified with reasons-responsiveness, or, more exactly, with having reasons-responsive mechanisms. This would at least be difficult as long as we accept that executing an intention or acting for a reason is sufficient for exercising voluntary control.¹⁹ The compulsive hand-washer acts for a reason and executes his intention to wash his dirty hands in doing so. However, as long as he would wash his hands even if they were not dirty (or he lacks another good reason to do so), we would not regard him as responsive to reasons in a sense strong enough for guidance control.²⁰ Furthermore, it seems that children often lack the capacity to *recognize* certain reasons, which is necessary for reasons-responsiveness (next to the ability to *react* to the recognized reasons). Yet, we will still regard them as executing their intentions or as acting for a reason in cases where they would not have recognized reasons for alternative courses of action among a suitable set of counterfactual scenarios where such reasons are present. As long as the execution of an intention, or

¹⁸ Even though their freedom may be very limited or, as McHugh says, ‘shallow’ (McHugh 2014a, 15), because they are not yet free ‘enough’ to be responsible (I thus agree with McHugh who explains that having voluntary control is not sufficient for responsibility, at least not in the sense in which children can already exercise voluntary control). Yet children’s actions exhibit freedom, intentionality and purposefulness different from ‘acts’ of stones, plants and maybe ‘lower’ animals (Hyman 2015 argues that there is a sense of ‘action’ in which also inanimate things can act).

¹⁹ McHugh 2014b equates voluntary control with executing intentions.

²⁰ Maybe he would not wash his hands anymore if he suffered extreme pain when doing so. Fischer and Ravizza distinguished between moderate and strong reasons-responsiveness to deal with such cases (1998, chs. 2 and 3).

acting intentionally, is not sufficient for the capacity of being reasons-responsive, we have to admit that there are cases of free action which cannot be completely captured by the concept of reasons-responsiveness, and consequently, not by the concept of guidance control.²¹

Now it follows, however, that since voluntary control is different from guidance control, any account that wants to explain our responsibility for actions solely by guidance control will remain *incomplete*. We need also an account of the nature of voluntary control, and of how voluntary control and guidance control play together in explaining our responsibility. Fischer and Ravizza do not provide us with such an account. They do not think that there is any more to our responsibility-grounding freedom than guidance control. I argued that this seems wrong.²²

I have now supported premise (2) of the main argument of this subchapter. We have seen that, given the existence of guidance control, we have two different kinds of control over our actions. Since we exercise both voluntary control as well as guidance control over actions, we can now ask which one of the two explains why our actions are subject to practical norms, or at least how both kinds of control play together in explaining action's subjection to practical norms. I have no clear idea how to answer this question. What I have shown, however, is that it is not clear that guidance control can explain subjection to practical norms by itself. Voluntary control also seems to be relevant for this task, given that we exercise voluntary control with respect to all our actions that are done for reasons or that can be understood as executions of intentions. Thus, it seems, according to the present argument, that believing for practical reasons is not possible just because we have guidance control over some beliefs. It seems that we also need voluntary control over belief.

9.5 SUMMARY

This chapter has considered the version of voluntarism that claims that we can just *believe, desire, feel, or intend* for practical reasons (rather than *form* attitudes for practical reasons). I

²¹ Even McHugh, who thinks that guidance control over our mind is more fundamental than voluntary control over our actions, has to admit that voluntary control over an action will imply that we are free in some 'shallow' sense, which does not underwrite our 'full' responsibility (McHugh 2014a, 9, 15) – but this just means that we are free in *some* sense when acting intentionally without our actions resulting from reasons-responsive mechanisms.

²² One might conclude from this that Fischer's and Ravizza's compatibilism is not complete, for even if it is clear that guidance control is compatible with determinism, now the question remains of whether voluntary control is compatible with determinism. However, their compatibilism is also motivated by Frankfurt-style counter-examples to the principle of alternative possibilities (PAP), which support their idea that responsible agency is not threatened by determinism – independently of what kinds of control responsible agency involves.

have argued that if this claim is meant to be supported by the idea that attitudes are ontologically similar to actions, then it is not well-supported, because attitudes do not admit of a categorization as ‘active’ due to their explanatory role in explaining behavior and due to the very concept of an attitude we are interested in within this investigation – as a state a subject is in for which the subject still seems to be directly responsible for (9.1). I then considered an objection to the effect that there is a category of ‘active states,’ which provided me with some opportunity for an excursion into my general picture of mental agency (9.2). I argued that introducing the category of an ‘active state’ is not meant to underwrite voluntarism and thereby pragmatism, but rather to make sense of epistemic normativity and other forms of rationality. I pointed out that according to my own conception of mental agency, we only have indirect control over our attitudes, including our intentions and decisions (9.3). Finally, I have argued, throughout my discussion of McCormick’s guidance control approach, that if the voluntarist’s position is meant to claim that we can believe, desire, feel, or intend for practical reasons *even though attitudes are not themselves actions*, then this version of voluntarism faces two problems (9.4). First, things that are as ontologically different as actions and attitudes cannot be controlled in the same way insofar as both kinds of things do not result from the mechanisms that ‘produce’ them in the same way. Secondly, guidance control seems not capable on its own to underwrite subjection to practical norms. This is because voluntary control also seems to play a crucial role, and seems to be different in nature from guidance control. This shows that the strategy of introducing a non-voluntary form of control over attitudes in order to vindicate voluntarism seems not to be very promising, because it is unclear how this non-voluntary form of control can explain why attitudes are subject to practical norms. Voluntary control seems to be relevant for this task as well, given that we exercise it over most of our actions.

Summary and Conclusion of Part III

Part III has argued that pragmatism implies voluntarism (chapter 7), but that we can neither form attitudes for practical reasons (chapter 8), nor just believe, desire, feel, intend for practical reasons (chapter 9). Our inability is not merely contingent. Rather, it is due to the ontological nature of attitudes as *states*. I have argued that strong mental voluntarists face a dilemma when they claim that we can *form* attitudes for practical reasons: either the forming of the attitude amounts only to indirect control, or it amounts to no control at all – depending on how we understand the notion of ‘forming an attitude.’ The only way out of the dilemma was to give up what lead us into it: that in order to believe, desire, feel, or intend at will, we have to be able to *form* these attitudes for practical reasons. According to this reply by the voluntarists, believing, desiring, feeling, or intending at will just means to believe, desire, feel, or intend for those reasons (rather than to *form* them for those reasons). This move tries to locate voluntary mental agency in the attitudes themselves. I have argued that this is impossible due to the passive element necessarily present in having an attitude. What we want to know is how we are responsible for this passive element of the attitude, rather than for activities that are supposed to make up the attitude according to the objector. It is this passive element that is unresponsive to our will. McCormick’s appeal to guidance control over belief did not help to save the idea that beliefs are subject to the same norms as actions are. Attitudes result from reasons-responsive mechanisms quite differently than actions do. This is not surprising if they are ontologically different kinds of things. Furthermore, the notion of guidance control seems incapable of explaining by itself why actions or attitudes are subject to practical norms – it is voluntary control that is required for this task.

Matthias Steup charges philosophers who understand the voluntariness of belief merely in terms of the capacity to form beliefs for practical reasons of ‘practical reasons chauvinism’ (Steup 2008, 387-388). The point of this criticism is that there might be concepts of voluntariness or doxastic freedom which consist in the capacity to respond to other kinds of reasons, like epistemic reasons. Am I guilty of practical reasons chauvinism? The notion of ‘voluntariness’ Steup and others have in mind is different from the notion I have examined here. It is about a kind of mental control that is *different* from the kind of control we exercise when we act for reasons. There is no reason for me here to deny that we might make sense of a talk of ‘control’ we exercise *in* believing, desiring, feeling, or intending, as long as we acknowledge

that it is not the same kind of control that we exercise when we act (though, as I have explained in 9.2 and 9.3, I have my suspicions about such a concept of control). If saying that we control our attitudes is just to say that they are responsive to reasons, then I am fine with it. As will become clear in the next part of the book, I will even grant that the reasons-responsiveness of attitudes gives rise to a specific form of direct responsibility for our attitudes. This responsibility will provide the basis for understanding the normative force of reasons for attitudes. It will allow us to meet the challenge of responding to the problem of mental responsibility – even though it will ultimately leave us with further questions and another challenge which I cannot fully meet within the present inquiry (the challenge of providing a full account of the normative force of the standards of rationality, i.e., the normative force of object-given reasons). It is time now to focus our attention on understanding my distinction between standards of rationality and practical requirements by thinking about how we are responsible for compliance with both, and thereby solving the problem of mental responsibility.

Part IV

Responsibility for Attitudes

Introduction to Part IV

I have argued at the end of part II that we should acknowledge a distinction between standards of the rationality of attitudes and practical requirements of action in order to capture the distinctive normative force that is attached to the latter, but not to the former. This normative force manifests itself in the appropriateness of passionate blaming responses towards people who fail to make an effort of will they owe to others – blaming responses like resentment, indignation, and guilt. The purpose of the present part is to spell out the distinction between standards of attitudinal rationality and practical requirements of action more clearly by connecting them to the relevant kinds of responsibility and blame that define them, and thereby providing a solution to the problem of mental responsibility.

My discussion is again structured in three chapters. Chapter 10 elaborates on a distinction between two forms of responsibility for attitudes that is common in the current philosophical discussion: historical responsibility for attitudes and answerability for one's attitudes. I argue that these forms of responsibility are genuinely *distinct* from one another. I show how this distinctness makes it doubtful that answerability is a form of responsibility, but how we can react to this doubt by acknowledging the two meanings of 'ought' distinguished at the end of part II in the 'Summary and Conclusion of Part II.'²⁸⁵

Chapter 11 then defends the idea that answerability is a form of responsibility with the argument from apology: we sometimes owe an apology to another person merely in virtue of our failure to properly respect them and thus also independently of whether showing respect was under our voluntary control; that we owe such an apology implies that answerability is a form of responsibility. We are thus, in a sense, directly responsible for our attitudes in the sense of being directly answerable for them, i.e., we are directly responsible for (non-)compliance with object-given reasons for attitudes. I spell out the two faces of responsibility for attitudes by associating them with the relevant blaming responses.

Distinguishing between these two faces of responsibility concludes with my *hybrid account of responsibility for attitudes*. *The easy solution* is true if we talk about the kind of

²⁸⁵ The claim that answerability is no form of responsibility is the equivalent to mental nihilism if we assume that there are reasons for attitudes only if we are (at least sometimes) responsible for (non-)compliance with our object-given reasons – a claim that is implied by the first premise of the argument from blameworthiness (cf. chs. 4 and 5). For denying that answerability is a form of responsibility is to deny that we are responsible for such (non-)compliance.

responsibility for attitudes which is associated with the passionate blaming responses. In order for it to be appropriate to blame someone passionately (later defined as *historical* blame: blame_H), we need to be in indirect control of our attitude. In this sense, our responsibility for attitudes – historical responsibility – is completely analogous to our responsibility for consequences of our actions. Yet in order to understand how we can be answerable for our attitudes, we need not assume that attitudes are under our direct control. Rather, it is sufficient to understand that they are responsive to object-given reasons: reasons that have a specific normative force insofar as whether we comply with them will affect our interpersonal relationships in specific ways. Blaming someone in the way that merely presupposes answerability (later defined as: blame_A) does not require anything beyond an attitude's reasons-responsiveness. We can thus rest content as involuntarists with respect to this face of responsibility.²⁸⁶

My solution to the problem of mental responsibility is thus a combination of indirect voluntarism and involuntarism, and I take each position with respect to another face of attitudinal responsibility, thereby avoiding self-contradiction.

Chapter 12 will react to an objection to my involuntarist position with respect to answerability: why am I so content that responsiveness to object-given reasons does not imply that we exercise direct, non-voluntary control over attitudes *in some sense*? My answer will be, first, that I am not content with this, but that involuntarism does not commit me to such a contentment. Rather, involuntarism is a claim about *how we can understand* one face of mental responsibility: it claims that we merely need to acknowledge that our attitudes are responsive to reasons which have a specific normative force. It does not require us to decide whether the reasons-responsiveness of an attitude somehow makes that attitude 'active.' Chapter 12 will defend the claim that we need not assume a form of direct, non-voluntary control that we exercise in believing, desiring, feeling, or intending in order to make it intelligible that we exercise a special form of mental agency over our own mind. I argue that we have special kind of *indirect* control over our own mind that we cannot – for conceptual reasons – exercise over the mind of others. This special kind of indirect control consists in rationally determining our attitudes by the activity of reasoning. Chapter 12 concludes that neither the problem of mental

²⁸⁶ Some authors think that reasons-responsiveness essentially involves some kind of mental agency that has a quite different nature than voluntary control (see esp. Hieronymi ms). As I explain below, I will remain neutral with respect to this claim. I do not think that we need to decide the question of whether a reasons-responsive state is essentially active in some sense in order to understand how we can be responsible for it. Knowing that it is responsive to reasons is enough for knowing that we are answerable and thus responsible for it. I return to the issue of whether reasons-responsiveness implies control at the end of chapter 11 and present it as an objection to my position. Part of my reply to this objection is to be found there, the other part is to be found in chapter 12, which I outline below.

responsibility, nor our intuitions about mental agency with respect to our own attitudes can by themselves motivate the idea of direct non-voluntary control over our mind. This leaves room for other motivations, maybe stemming from the very idea of agency (cf. Hieronymi ms).

Why, one might wonder, did I not start right off with part IV after part I? Why did I need all the discussions about reasons and rationality in parts II and III? Part II was essential because I have argued, in part I, that the problem of mental responsibility is a problem about understanding how we are responsible to the standards of rationality. Part II was then concerned with the status of these standards. I spelled out arguments that showed us that the very *existence* of reasons for attitudes provided by these standards can be reasonably called into doubt. By this, I motivated the project of spelling out the normative force of reasons for attitudes. Yet I concluded, by giving some initial replies to the arguments, that we should instead call into doubt the status of the standards of rationality as requirements similar to the practical requirements of action. By this, I provided the framework for my reply to the problem of mental responsibility. The present part will give more substance to the distinction between standards of rationality and practical requirements by connecting them to different forms of responsibility as well as associated forms of blame.

Part III showed us that we cannot account for normative force of reasons for attitudes by revising the understanding of reasons and rationality I presupposed in part II, i.e., by understanding attitudinal rationality as a matter of responding correctly to practical reasons for attitudes. I argued that pragmatism implies believing at will, and that we cannot believe at will. While part II excludes mental nihilism as a reply to the problem of mental responsibility, part III excludes mental pragmatism as accounts of the normative force of reasons for attitudes. Part II in addition sketched the framework of my solution. Part IV now provides more details to my solution.

One might not agree with some of these details of part IV. Some people balk at the idea that they can never appropriately *resent* another person for having a hostile attitude towards them without assuming that the person had indirect control over their attitude. I will provide arguments for this idea. Yet if one does not agree, part II can also be read as providing an independent motivation for making a distinction between two faces of responsibility. Let me explain.

Note first that there will be two faces of responsibility as long as there is a plausible distinction between

- (a) a set of responses R_1 where the appropriateness of the responses in R_1 presupposes that the person at which the responses of R_1 are directed had indirect voluntary control over their attitude, and
- (b) a set of responses R_2 , where the appropriateness of the responses in R_2 does not presuppose indirect voluntary control.

Whether the line between the two faces of responsibility is the line I draw in this part – i.e., between passionate and non-passionate responses – does not matter for whether there is such a distinction between R_1 and R_2 .

Next, remember that the conclusion of part II was that we have to introduce two kinds of ‘oughts’ and reasons in order to be in a position to respond to the arguments from blameworthiness and from cases of conflict. In chapter 10, I will present another argument that underwrites this ambiguity in ‘ought’ and ‘reasons’ (see especially 10.4). In one sense of the terms, that someone ought to have an attitude or has a reason for an attitude presupposes voluntary control, while in the other sense it does not. As long as we are sometimes appropriately reacting to the violation of both ‘oughts’ with two different sets of reactive attitudes, each of which presupposes that we are responsible (in a sense), we have a distinction between two faces of responsibility.

Thus, part II already provided the material to distinguish between a form of responsibility that presupposes indirect control, and a form of responsibility that does not presuppose control. If one does not agree with me about which kinds of blaming-responses to associate with each type of responsibility for attitudes, one can feel free to draw the line somewhere else, but nevertheless accept the general structure of my solution to the problem of mental responsibility.

Thus, part II played an essential role in my overall argument: it motivated our search for a solution to the problem of mental responsibility by considering the normative force of reasons for attitudes; it provided the framework for my reply to this problem; together with part III, part II excluded mental nihilism and pragmatism as possible accounts of the (lack of the) normative force of reasons for attitudes, and thus left us only with the option of looking for a solution to the problem of mental responsibility while assuming a mental constitutivist account; finally, part II backs up my framework of the constitutivist account if the details I spell out in part IV would turn out to be false. I return to this backup-account in the conclusion of the book once more.

Let us now turn to our task of understanding the two faces of mental responsibility.

Chapter 10

Historical Responsibility and Answerability

I begin this chapter by briefly re-stating the central problem of this book and my reply to it (10.1). I then introduce two (putative) kinds of responsibility for attitudes – indirect responsibility and answerability – by distinguishing two kinds of blameworthiness for attitudes (10.2). I argue that these two kinds of blameworthiness cannot be compared with each other in such a way as to reach judgment about the overall blameworthiness of a person for an attitude (10.3). By putting my overall argument in a syllogism, I show that this conclusion is problematic only if we assume that ‘ought’ means the same when applied to attitudes that are indirectly controlled and attitudes that are not indirectly controlled (10.4). Distinguishing two faces of normativity thus allows us to acknowledge two faces of responsibility. Chapter 11 will turn to the details of these two faces.

10.1 THE ORIGINAL PROBLEM RE-STATED

Philosophical thought about responsibility is structured by taking *actions* and their *consequences* as the kinds of things for which we are responsible.¹ This way of thinking about responsibility highlights a difference between two modes of being responsible: direct and indirect responsibility. On the one hand, our actions are exercises of voluntary control, and we perform our actions for reasons. Their causal consequences, on the other hand, are not exercises of control, and they are no things performed for reasons.² Our responsibility for consequences

¹ First, take one of the central questions of normative ethics: what are the right-making features of an *action* (e.g., its consequences, motives, or properties of the action itself)? Secondly, take the debate about free will and determinism, where we ask whether we are free with respect to and thus responsible for our actions in a way that is compatible with the causal structure of the universe. Thinking about our responsibility for and control over *attitudes* (instead of actions and their consequences) might well give us a clue as to how to solve this traditional problem (cf. Hieronymi ms; Wagner 2015). However, I will not be concerned with the connection of my claims to these traditional discussions here.

² I thus exclude other actions from the consequences of an action. Consequences in my sense are *mere* consequences which are caused by previous actions, but which are not themselves directly controlled. It is important for my account that I include attitudes in the class of potential consequences of actions. I do this by saying that attitudes are no things *performed* for reasons (although one might say that they are *held, formed, or even done* for reasons).

originates in our responsibility for actions which cause them.³ This allows us to say that we are directly responsible for our actions, but only indirectly responsible for their consequences. Whenever we are responsible for a consequence, our responsibility can be *traced back* to our responsibility for prior actions.⁴

If we try to put attitudes within this traditional framework, we are faced with the puzzle that is central to this book (see chapters 0.1 and 2). Let me restate this puzzle by highlighting two plausible lines of thought which result in claims that are in conflict each other. The task of the present and the next chapter will be to reconcile the two claims.

On the one hand, it seems that we do not have direct control over our attitudes. In this respect, attitudes behave similarly to mere consequences of our actions. Indeed, it seems undeniable that they are *sometimes* consequences of our actions, and that we thus have, to a certain extent, *indirect* control over them: we can manage our emotions through meditation, and we can form justified beliefs by proper investigation. Yet it is hard to see what another, *more direct*, kind of control over attitudes is supposed to look like. If we lack any direct mental control, it seems that we should conclude that we are *only* indirectly responsible for our attitudes – in the way we are responsible for consequences of our actions. This line of thought motivates *The easy solution*.

On the other hand, our attitudes are within the ‘space of reasons’: they are subject to evaluations to which brute consequences of actions could never be subject to. Attitudes cannot only be evaluated as better or worse to have (as consequences can be evaluated as better or worse). Rather, we think of our attitudes as being rational or irrational. In this respect, it seems that our attitudes do not behave like mere consequences of our actions. Given their presence in the space of reasons, they seem to be much more like actions rather than their consequences. We are tempted to conclude from this second line of thought that we must be *directly* responsible for our attitudes. This line of thought motivates a denial of *The easy solution*.

Luckily, we need not decide between both lines of thought. We can endorse both. We can claim that each of them is sound as long as we say that each of them is sound for another face of responsibility for attitudes. I will now turn to providing the contours of the two faces by distinguishing between historical responsibility and direct answerability – two concepts that are prevalent in contemporary theories of responsibility for attitudes. I highlight the distinctness

³ I do not claim that we are responsible for all the consequences of our actions, of course. The point is rather that if we are responsible for consequences of our actions, then we are so in virtue of the fact that we are responsible for actions which caused them.

⁴ Cf. Fischer/Tognazzini 2009 on tracing.

by describing cases in which we are blameworthy in a sense that presupposes one face of responsibility, but not blameworthy in a sense that presupposes the other face of responsibility.

10.2 TWO KINDS OF (PUTATIVE) BLAMEWORTHINESS

Per stipulation, the person in the following example has no reasonable opportunity to avoid their irrational attitude-response of becoming angry:

Caring Todd

Todd was mildly, or not at all, offended, and reacts with an inappropriate degree of anger. Intuitively, we might judge that Todd is blameworthy. His reaction is irrational and harmful to another person.

However, we learn that Todd took anger-management classes for years, and does various exercises in his daily life, which were recommended to him by specialists. Whenever he misses out a class, and does not do his exercises, he will occasionally show an irrational degree of anger. He knows about this tendency, but there is now further way for him to control these angry impulses.

However, this week, Todd had to take care of his sick mother. Let us assume that this took up almost his whole time and energy, but it was nevertheless his moral duty. As a result, he could not do his exercises or visit his class without neglecting his duty. Any further efforts we could reasonably expect him to make to avoid his anger-response were ineffective.

Compare the case of Todd with the following case:

Devastation-Case

Tony offends Tom. Tom could now get angry, and Tom's anger would be an appropriate reaction. However, if Tom gets angry, then Tony would notice Tom's anger and feel terribly devastated. Tom knows this and has the ability to manage his anger (e.g., by taking a deep breath). However, Tom omits this action of managing his anger and gets angry. Tony ends up being terribly devastated. Let us assume that, all things considered, it would be better if Tony did not feel terribly devastated, and Tom knows this normative fact about the situation.

The difference between Todd and Tom is that they seem to be able to justify their attitudes in different ways. While Todd can offer reasons for neglecting his anger-management, and thus

can *historically excuse* himself for being angry, Tom cannot do so. This is because Tom *had* a reasonable opportunity to manage his anger but did not make use of it. Thus, Tom is not historically excused in the way Todd is.

However, Tom might justify his anger in *another sense*. For Tom's anger is a rational reaction to the situation. Tom seems thus to be justified in what I will call the *answerability-sense* of responsibility. Tom can justify his anger by giving object-given reasons for his anger. Todd, in contrast, cannot do so, and is thus not justified in his anger in the answerability-sense.

This allows us, *prima facie*, to distinguish between *two faces of blameworthiness*: the direct blameworthiness_A for attitudes (answerability-blame) and the indirect blameworthiness_H for attitudes (the blame of historical responsibility) which can be distinguished from each other by the following conditions which are necessary for one kind of blameworthiness, but not for the other:

Blameworthiness_A: *S* is *blameworthy_A* for their anger only if *S*'s anger is irrational given the offense. (i.e., only if it is insufficiently supported by object-given reasons for the anger.)

Blameworthiness_H: *S* is *blameworthy_H* for their anger only if *S* did not perform an action *S* ought to perform and which might have avoided or mitigated the anger. (i.e., only if *S*'s omission is insufficiently supported by reasons.)⁵

Given these definitions, Todd is blameworthy_A, but not blameworthy_H, whereas Tom is blameworthy_H, but not blameworthy_A.

Thus, these cases help us to keep apart two faces of responsibility or blameworthiness which are currently distinguished within the literature. Especially Pamela Hieronymi and Angela Smith, who have recently elaborated on two (similar) accounts of responsibility which locate direct responsibility (also) in attitudes, seem to be committed to these two different ways of being blameworthy for attitudes.⁶ This opens up the space for both indirect responsibility for attitudes (due to our direct responsibility for actions by means of which we sometimes cause our attitudes) as well as direct responsibility for attitudes (due to our answerability for reasons-responsive attitudes): If you are blameworthy_A, then this means that certain forms of criticism

⁵ Of course, *S* might also be blameworthy in this historical sense if *S* *did* perform an action *S* ought not to have performed, so that the blame would be appropriate due to an action rather an omission of *S*. I omit this for simplicity.

⁶ As I read them, they do not wish to exclude actions from the sphere of direct responsibility, but rather argue that the conditions necessary and sufficient for direct responsibility are satisfied by some actions as well as by some mental states. That her account is meant to also explain direct responsibility for intentional actions is explicitly pointed out by Hieronymi (2014, 15-16).

or blame are appropriate which are associated with what Smith calls *answerability* for attitudes (i.e., direct responsibility for attitudes). If you are blameworthy_H, then this means that certain forms of blame (e.g., moral blame for failing to properly manage your mind in the way you owe it to others) are appropriate that are associated with historical responsibility for attitudes.

According to Smith, both kinds of responsibility are, fundamentally, matters of *being answerable*. If you are responsible for something, then you are in the position to *respond* to requests of justification (Smith 2005; 2015b). The requests of justification are then *appropriate*, or *intelligible*, or *correct*. Smith argues that this normative property of the relevant requests is clearly given in the case of most intentional mental states: we can ask people for their reasons for believing that there will be nice weather tomorrow, for their reasons for intending to go to the concert, or for their reasons for being angry about this silly remark. When we do this, we do not merely request a *causal* explanation of how those mental states came into existence. Nor do we ask for the reasons the people might have had to bring themselves into the relevant states by previous actions. Rather, we are asking for reasons which *rationalize*, or *justify*, the mental states in question. These requests are appropriate because the mental states reflect – or are supposed to reflect – our rational judgment about what is true, worthwhile, right, and so on (Smith 2005, 250-264; 2015b, 103-104).

Hieronymi similarly argues that in having a mental state (as well as in acting), we ordinarily reveal our answer to a certain question (Hieronymi 2006, 53/54, 56; 2014, 12-17): in believing that *p* we reveal our answer to the question of whether *p* is true; in intending to do *A* we reveal our answer to the question of whether to do *A*;⁷ in desiring *x* we reveal our answer to the question of whether *x* is good (in at least one respect); in fearing *x* we reveal our answer to the question of whether *x* is dangerous. We reveal, as Hieronymi claims, an important piece of our mind, or the *quality of our will* (Hieronymi 2014, 15). Thus, those mental states by their very nature imply that we are answerable for having them, that a request for justification is appropriate, intelligible, or correct. In contrast to mere headaches, or in contrast to implanted attitudes, they tell us something important about ourselves, about our overall outlook on the world and our place within it. This sounds similar to Smith's condition of judgment-reflection.

Both Smith and Hieronymi thus claim that there is no need to suppose voluntary control over mental states in order to explain why we are responsible for them. Rather, it is rational control by judgment (Smith) or evaluative control by 'answering a question' (Hieronymi) that grounds responsibility. I will now argue that Hieronymi's and Smith's accounts imply *two*

⁷ The same holds for *intentionally doing A* (Hieronymi 2014, 15), which implies having the relevant intention.

distinct forms of responsibility. I will show in the subsequent section (10.4), how this can give rise to a problem if we do not assume that ‘ought’ is ambiguous between a sense presupposing voluntary control and a sense that does not presuppose control.

10.3 DISTINCT KINDS OF (PUTATIVE) RESPONSIBILITY

In order to see how accounts like Smith’s and Hieronymi’s imply two distinct forms of responsibility, consider once again the case of responsibility for having a headache. We could be responsible for our headache insofar as we caused it by previous actions and could, at the time of the actions, be expected to foresee that the headache will occur. Thus, I might have omitted to take a pill which would have prevented the headache. We can say, with Hieronymi’s terminology, that I had *managerial* control (that is, indirect voluntary control) over my headache, because I had the ability to manage whether I will have it by attending to the regularities of the physical world and using them for my aims (i.e., by taking a pill):

We exercise this sort of control when we manipulate some ordinary object to accord with our thoughts about it. We typically control ordinary objects by performing intentional actions which affect that object in the way we intend. Of course, our control over such objects is never unlimited. We are subject to physical and temporal limitations, to limitations of skill, and to luck. Importantly, the degree to which we exercise control over some object is measured not by the absence of such limitations – as though we would exercise greater control over our coffee cups if they did not obey the laws of physics – but rather by our ability to navigate, manage, and make use of those limitations in order to accomplish our purposes. In fact, in many cases, exactly those features that seem to limit our control also make such control possible. We can control ordinary objects at all only because we know they observe certain regularities (Hieronymi 2006, 53).

We have the same kind of control over our mental states. We can learn the regularities by which we come to have, for example, anger, and then set out to manage this anger by, for example, avoiding certain situations, or engaging in meditation. The fact that we can manage our anger through actions makes us *indirectly* responsible for our anger. Or, to put it in Hieronymi’s terminology, the anger *falls in our jurisdiction*.

It is important to see, however, that this is not the only sense in which we can be responsible for our anger if we adopt Hieronymi’s (or, for that matter, Smith’s) conception of responsibility. We are also *directly* responsible for our anger insofar as it embodies, or reveals, our answer to the question of whether there are object-given reasons to be angry (for example, our *judgment* that somebody offended us). The fact that our anger falls into our jurisdiction

makes us responsible for our anger in one sense; the fact that it embodies our answer to the relevant question, or reveals our judgment about object-given reasons, makes us responsible in another sense:

Our responsibility for and agency with respect to the relevant actions and attitudes thus have two distinguishable aspects: we are answerable for them, insofar as they embody our answer to certain questions, and they fall into our jurisdiction, insofar as we are expected to manage and control them through our actions (Hieronymi 2014, 24).

It is important to see that the two kinds of responsibility involved here are *distinct*, i.e., none of them is reducible to the other. Our answers to requests of justification will, depending on the *kind* of request, entail different kinds of reasons: when asked why I failed to manage my anger, I can only justify myself by citing reasons for action; when asked why I am angry, I will ordinarily justify myself by citing object-given reasons.⁸ The first kind of answer could contain my reasons for my omission of failed anger-management. I might, for example, say that I had better things to do than to avoid the relevant situation where I knew I would feel angry (my mother was sick, I had to take care of her, and thus I had no time to attend to avoiding those situations). By citing reasons for actions which lead to the relevant state of anger, I could excuse myself in some sense. The second kind of answer, however, might include mentioning that I was badly offended, and might thus rationalize my anger by giving object-given reasons. This answer can also excuse me for my angry mood if I had in fact good reasons to feel offended.

Introducing these two concepts of responsibility will lead to cases in which it is far from clear how we are supposed to evaluate the overall blame- or praiseworthiness of a person. If both kinds of responsibility are *genuine*, one would expect that certain reactive attitudes – forms of praise and blame – will be appropriate when we fail to justify ourselves by giving the answers associated with the respective kind of responsibility. If this is the case, however, we will end up with scenarios where a person is blameworthy in one sense, but not in another. I gave examples of such cases at the beginning of subchapter 10.2 (*Caring Todd* and *Devastation-Case*). Take also, for further illustration, the following case:

⁸ I say ‘ordinarily’ because we saw in ch. 7.1 that statements about what we ought to believe, desire, feel, or intend can be understood as questions about my reasons for bringing me into or maintaining my mental state.

Ronja the racist

Ronja was raised in a racist community. The people she lived with regularly made racist remarks and did not allow opposing opinions to co-exist within their community. Furthermore, Ronja never had any contact to people outside her community and had barely access to general education. As a result of her unfortunate history, Ronja holds racist attitudes.

How would the answerability-theorist describe this case? First, they would point out that since Ronja holds a full-blooded racist attitude, she is answerable for it: we can request her reasons. If she cannot give an adequate justification for it, she is *blameworthy_A*. Yet, secondly, Ronja could not have been reasonably expected to manage her disrespectful attitude. She is not blameworthy for any actions that led her to this attitude (at least she is not blameworthy for these actions in virtue of the fact that they led to her racist attitude). It seems that answerability-theorists have to admit that she is *blameless_H* for her racism.

(How) Are we supposed to weigh these two judgments? While Hieronymi thinks that her account adequately captures the evaluative ambiguity involved in cases like *Ronja the Racist*, it can seem that this ambiguity is not a satisfying result. Rather, we want to know whether Ronja is blameworthy or blameless, *all things considered*. Ronja either ought not to have her racist attitude or she is allowed to have it. Saying that she is blameworthy in one sense but not in another seems to amount to saying that she is allowed to have the attitude in one sense, but ought not to have it in another. This seems, *prima facie*, to be an odd result.

It is important to note that this evaluative ambiguity in certain cases is not restricted to Hieronymi's and Smith's positions. Rather, it applies to *all* accounts which assume that we can be directly responsible for our mental states, although we lack direct voluntary control over them.⁹ As soon as we separate responsibility from voluntary control, direct responsibility will no longer be a responsibility *to practical norms*. Having voluntary control over X means to be able to do or have X for practical reasons in the way we can perform actions for such reasons. Our *exercises* of voluntary control (i.e., our actions), and *only* those exercises, are thus subject to practical norms. We can be criticized for failing to comply with practical requirements of prudence and morality if we fail to respond correctly to our practical reasons for action. If there is a kind of blameworthiness that is completely *independent* of voluntary control, this would

⁹ This includes accounts like those of Adams 1985, and Owens 2000, who argue that responsibility for mental states is not a matter of control at all (except for intentions, which are reflectively controlled according to Owens). For Owen's overall account, a distinction between direct responsibility for belief to epistemic norms, and indirect responsibility for belief to practical norms (i.e., responsibility for the actions which cause beliefs) is essential. The main focus of his work is on the direct responsibility we have for our beliefs and other attitudes.

thus imply that this blameworthiness is *sui generis*, that is, it is not made appropriate by failures within the domains of prudence or morality. Our failure would not be that we did not react correctly to reasons for actions. Rather, our mistake would be that we did not react correctly to object-given reasons for attitudes. If we were blameworthy for both kinds of mistake, this would imply two kinds of responsibility for attitudes that cannot be reduced to one another – indirect responsibility to practical norms of action (which cause attitudes), and direct responsibility to the standards of the rationality of attitudes.

Before we spell out this problem for the idea that we are directly answerable and responsible for our attitudes, we should note an important ambiguity in the case of *Ronja the Racist*. We might imagine Ronja to be fully rational: her racist attitudes would then be justified from her own distorted perspective.¹⁰ Yet we might think that she is blameworthy_A because she is unable to provide a justification acceptable *from our perspective*. This would have two philosophically interesting implications: First, it would imply that someone who is *fully rational* can be blameworthy. Secondly, it would imply that being blameworthy_A is not an absolute property, but rather a property *relative to the perspective of the person who blames*. If someone has an attitude that is not justifiable from my perspective, I would appropriately regard the person as blameworthy_A. Yet for you, the person might not be blameworthy_A because, given your system of beliefs, the justification the person offers from their perspective might be sound.

How plausible are these implications? A problem for the first implication is that it is *prima facie* hard to see why someone should be blameworthy if they responded correctly to *all* their reasons. Being blameworthy or criticizable in the sense at issue seems to presuppose that we failed in our *own* lights – i.e., failed to comply with the reasons available to *us*. A problem for the second implication is that everyone would be blameworthy for their attitudes given *some* perspective. There will always be a system of beliefs conceivable from which you would be blameworthy_A.

These implications might not turn out to be as troublesome if we pay attention to the nature of the responses that make up blame_A, and to which I return in the next chapter in some more detail. For it might seem that distrusting Ronja is appropriate independently of whether her attitude is rational or not. It might be inappropriate *for someone within her community* to distrust her, and yet *we* might judge that it is *appropriate for everyone* to distrust her. We regard her attitudes as disrespectful and dangerous, and this will affect our overall outlook and judgment of how both we and others should relate to Ronja. Yet, of course, the person who shares

¹⁰ On the ‘rational racist,’ cf. Basu 2019.

Ronja's perspective will not rationally regard her as having an objectionable attitude, and thus not rationally think that the relationship Ronja has towards other people should be impaired by her attitude.

I have doubts that the form of distrust we might show towards a *fully rational Ronja* is the same as the distrust we show towards the *irrational Ronja*. I have this doubt for two reasons. The first is that if there is such a thing in our world as a fully rational racist, this person might have a radically different set of beliefs (*Weltbild* – 'world-picture') from ours. In such a case, it might be hard or impossible for us to understand *what exactly* the person believes. In this case, we cannot regard them as being answerable for their beliefs, because we do not know what reasons would possibly justify their 'views.' However, and this is my second reason, even if we grant that we can understand the fully rational racist, we would have a tendency towards adopting Strawson's (1962) *objective stance* towards them: regarding them as a problem to be dealt with rather than as someone who is fully responsible. If our reasons for rejecting their view do not move them to re-think their racist attitude, then we would just have to deal with them in non-rational ways – for example, not by reasoning, but by ensuring in some way that they do not harm others or spread their attitude.

I will not decide here whether being blameworthy_A is possible even if one is fully rational, and is thus a property relative to the person who blames. I suspect that it is not. If it was, this would imply that blameworthiness_A is not grounded in a person's irrationality, but rather in whether the person holds an attitude that is objectionable or disrespectful given *our* perspective, rather than their own perspective. This would pose a problem for understanding the normative force of rationality in terms of blame. It would *not* pose a problem, however, for understanding the normative force of object-given reasons in terms of blame (i.e., it would not pose a problem for the normative force that matters for my investigation here). For, given our perspective, rational Ronja would have a view that is insufficiently supported by object-given reasons that are available *to us*, and she would be blameworthy_A in virtue of the fact that she holds this view that is insufficiently supported in *our* lights. It would merely be false that she is blameworthy_A in virtue of the fact that she did not comply with *her* reasons. Yet her blameworthiness_A would be understandable in terms of object-given reasons – more precisely, in terms of those reasons accessible to *us*.

Since I suspect, for the reasons stated above, that being blameworthy and criticizable (even in the sense required for answerability-blame) presupposes that one fails in one's *own* lights – i.e., fails to respond correctly to one's *own* reasons –, I will understand *Ronja the Racist* as a case involving irrationality. Even if Ronja was raised within a racist community, which

makes it understandable to us that she holds the racist attitudes she holds, we can imagine that she has sufficient evidence available to her to draw the conclusion that her attitude is unjustified. Yet she did not wrong anyone in her voluntary conduct – her irrationality is purely a result of external influences. We might conceive of her case as a case of ‘social implantation’ of an irrational attitude – i.e., as analogous in certain respects to the implantation of an attitude by science-fiction surgery.¹¹ She is thus still blameless_H. In what follows, I assume this version of the case. I will assume this understanding for all other cases I will discuss that share the ambiguity of *Ronja the Racist* between a case where the subject is irrational and where the subject is rational: I choose the reading of the case where the subject is irrational.

10.4 A PROBLEM FOR ANSWERABILITY AS A FORM OF RESPONSIBILITY?

If the two kinds of responsibility for attitudes – answerability and historical responsibility – are as distinct as I have presented them, then the idea that answerability is a genuine form of responsibility is faced with the problem I have sketched in the last subchapter immediately after first presenting *Ronja the Racist*. Here is the problem put in a syllogism. I use the *Devastation-Case* from 10.2 for the argument instead of the Ronja-case in order to avoid some complications:

- (1) If we are not merely indirectly responsible for our attitudes, but also directly responsible, then there are cases in which we are *blameworthy_H* but *blameless_A* for an attitude M (e.g., *Devastation-Case*).
- (2) If we are *blameworthy_H* but *blameless_A* for M, then we both ought not to have M and are allowed to have M.
- (3) It cannot both be true that we ought not to have M but are allowed to have M.
- (4) Thus (from (1), (2), and (3)), we are not both indirectly and directly responsible for M.
- (5) However, we are indirectly responsible for M.
- (6) Thus (from (4) and (5)), we are not directly responsible for M.

¹¹ An important difference is, however, that Ronja’s irrationality is part of her overall character that developed over her whole life while a literally implanted attitude can be like an alien intruder in one’s mind that occupies it suddenly. Yet I think that also in the latter case, we would hold the person responsible in a certain way for their attitude insofar as it is immediately after the implantation responsive to reasons: we can be answerable for implanted attitudes.

Since Tom (our protagonist in the *Devastation-Case*) is, per definition, directly answerable for his attitude – it is responsive to reasons –, it follows from (6) that being answerable is not being responsible.

I agree with (1) and (5). (1) just states how proponents of an answerability-account of direct responsibility for attitudes would have to analyze *Devastation-Case*, given my definitions of answerability-blame and historical blame. (5) is true because indirect control over attitudes is sufficient for being responsible for attitudes: there are at least some cases where we are indirectly responsible for M, and we can stipulate that we deal with such a case in the argument. Since I want to defend (6) under a certain reading, I have to deny either (2) or (3).

I do not think it is a promising strategy to argue against (2). There are two strategies one might try to employ against (2). The first is to adopt objectivism about reasons: whether you are blameworthy will depend on your epistemic position – on whether you knew or could have known what the right attitude was; but whether you ought to have the attitude does not depend on your epistemic position.¹² However, even if we are objectivists, we can easily modify (1) and (2) so as to make these premises uncontroversial:

(1*) If we are not merely indirectly responsible for our attitudes, but also directly responsible, then there are cases in which we are *blameworthy_H* but *blameless_A* for an attitude M *and in which we know or are in a position to know how M is supported by reasons and how any reasonable actions available to us that would cause or maintain M are supported by reasons.*

(2*) If we are *blameworthy_H* but *blameless_A* for M, *and we know or are in a position to know how M is supported by reasons and how any reasonable actions available to us that would cause or maintain M are supported by reasons*, then we both ought not to have M and are allowed to have M.

To illustrate a relevant case, imagine that in *Devastation-Case* Tom knows that his angry response would be sufficiently supported by object-given reasons, but that his omission of not taking a deep breath in order to avoid this response is insufficiently supported by reasons. In this case, Tom knows how both his attitude and his actions/omissions of attitudinal management are (un)supported by reasons. As pointed out earlier, he is thus *blameless_A* but *blameworthy_H* for his anger. It is hard to see how an objectivist about reasons can, *merely in virtue of being an objectivist*, deny that Tom, in a sense, both is allowed to become angry and ought

¹² On objectivism and subjectivism, cf. ch. 4.2.2.

not to become angry. In order to avoid this conclusion, the objectivist would have to argue that one of the kinds of blameworthiness/blamelessness involved are not genuine. But this would just confirm the conclusion of the argument (1)-(6) that answerability is not a genuine form of responsibility.¹³ I thus conclude that, by reformulating the premises in the way I suggest with (1*) and (2*), they become plausible also for objectivists.

Yet there might be other grounds for rejecting (2) than objectivism. The second strategy against (2) is to explain how to compare the judgments about the two kinds of blameworthiness with one another in order to reach a conclusion about what attitude the person ought to have, all things considered. To illustrate such a strategy, let us return to our case of Ronja. How could we compare the judgment that Ronja is blameworthy_A with the judgment that Ronja is blameless_H (remember that here it was the other way around as in the *Devastation-Case*) so as to reach a judgment about the overall hypological status (blameworthy/blameless) of her attitude (and thus about its deontic status)? Probably Smith can be read as employing this strategy when she argues that a criminal who was abused as a child

is fully answerable for, and accountable to us for, his crimes, but the fact that we as a moral community allowed him to suffer such terrible abuse as a child is part of the ‘answer’ he is likely to give when we demand that he justify his conduct; while such an excuse does not get him fully off the moral hook, it should make a difference to how we respond to him morally (Smith 2015b, 114-5).

The case of the abused criminal is, in a respect, similar to Ronja’s (or Todd’s) case: we can conceive of the case in such a way that the criminal had no reasonable opportunity to avoid his crime, as Ronja (or, for that matter, also Todd) had no reasonable opportunity avoid her irrational attitude. Given the criminal’s perspective (which he did not choose), it seemed justified to him to commit his crime. Yet it seems that, even if we understand the case in this way, we would not want to let him ‘fully off the hook,’ and this is not merely for pragmatic reasons (e.g., in order to deter others). Suppose his crime was a cruel murder. His actions, even though they might have seemed justified to the criminal himself at the time he committed the murder (‘I have a right to pay back to society what they have done to me’), were still genuine activities located within the space of reasons. Given the intentionality of his conduct, we still want to regard the criminal as blameworthy for his attitudes even if he lacked any opportunity not to commit his crime.

¹³ One could in principle also deny (5). Yet it is hard to see how one could deny that indirect responsibility for attitudes is a genuine form of responsibility: it is just the kind of responsibility we also have for consequences of our actions.

Up to this point, I can fully agree with Smith, and we do not yet have an objection against (2). We would get an objection to (2) if we were to interpret Smith as follows. Our responses to the blameless_H are different than our responses to the blameworthy_H.¹⁴ Yet the people which I call ‘the blameless_H’ are, according to Smith, not in fact blameless (‘in a sense’). Rather, we can judge them as *overall blameworthy* to a certain degree by acknowledging that they are still answerable for their attitude. This idea of ‘being blameworthy to a certain degree’ would commit Smith to comparing the two hypological judgments involved in cases like *Ronja the Racist*, *Caring Todd*, *Devastation Case*, or the victim criminal so as to come to an overall judgment about the degree of overall blameworthiness. It would at the same time avoid introducing two concepts of blameworthiness, as I propose we should do.

However, it is hard to see how such a comparing is supposed to work, because the two relevant judgments are based on very different kinds of reasons: that Ronja is blameworthy_A is based on her lack of object-given reasons for her racist *attitude*; and that she is blameless_H is based on her lack of reasons for *actions* which would have led her not to have the racist attitude. If we want to reach an overall hypological judgment, it is not even clear what kind of question we are supposed to ask in order to reach that judgment: What should Ronja *do*, all things considered? Or, what ought she to *believe*? Or, more generally, what kind of attitude ought she to have? Only reasons for *action* seem to be relevant for the first question, and for the latter two questions only reasons for *attitudes* seem to be relevant. There does not seem to be any question about how Ronja should *respond*, *all things considered*, where both reasons for action and reasons for attitudes go into the deliberation. For what is this ‘response’ supposed to be – an action or an attitude, or a hybrid of both?

Note that it will not help to propose a strategy of how to weigh object-given reasons against practical reasons for attitudes along the lines of McCormick (2017) and Reisner (2008). For these kinds of reasons are still both reasons *for attitudes*. The problem here is more severe: there is no deliberation of weighing reasons for different kinds of responses with each other, because it is unclear what one thereby wants to deliberate *about*.

Especially given my argumentation throughout part II, it seems to be more promising to look closer at premise (3). If we accept all other premises as well as the conclusion, then (3) is false. This is, I take it, further proof that we should acknowledge an ambiguity in ‘ought’ and ‘reasons’ (cf. ‘Summary and Conclusion of Part II’). Acknowledging this ambiguity saves

¹⁴ Note that I regard Ronja, Todd, as well as the victim criminal as blameless_H. I can do this because I merely stated a necessary condition on blamelessness_H above. I take victim criminals in the way I described the case as paradigm cases of people who are blameless in the historical sense of ‘blameless.’

the idea that we are directly responsible – i.e., answerable – for our attitudes. I turn now to how this direct answerability differs from indirect responsibility, thereby defending (3), and how acknowledging this difference solves the puzzle of mental responsibility. My elaboration of this difference will take up the rest of the present chapter as well as chapter 11. Let me begin by stressing the face of historical responsibility: the sense in which we are never directly responsible for failing to respond correctly to our reasons for attitudes.

10.5 HOW WE ARE *NOT* BLAMEWORTHY FOR IRRATIONALITY

Take first the face of responsibility that does not apply directly to our attitudes: there is a sense in which we are never blameworthy for an attitude if that attitude was outside of our indirect control. Consider the case of *Caring Todd* from 10.2 again. Todd could have, in principle, decided between two options: (a) doing what is right and having an irrational and harmful emotion; (b) doing what is wrong and avoiding such an emotion (by visiting his class, etc.). This seems like a dilemma. However, if we grant that it might well have been better for Todd to care about his mother (there was no one else who could do so, let us assume) than to avoid the irrational and harmful emotion, then option (a) was, objectively, the better option. Acting as best as he could, he ended up having an irrational and harmful attitude. Why, we might wonder, should he end up blameworthy for acting as best as he possibly could? Sure, the outcome is not optimal. But there just was no better outcome available, given human limitation (and, given the details of the case, we can even say ‘given conceptual limitation’).

Note that, depending on his capacities, Todd might well *become* indirectly responsible for his anger. Assume that his emotional reaction is not just brief but takes place over a period of time. It is plausible that at some point, Todd should take measurements to regain his posture. Insofar as he is not a mere victim of his feelings here (what I stipulate in order to rule out other explanations for his lack of responsibility, like a lack of reasons-responsiveness), he will have the capacity to do so at some point after the trigger of his anger took place. Insofar as Todd fails to take advantage of any reasonable opportunity to manage his anger-response, he is blameworthy for showing this response. Such an opportunity might, among others, be an opportunity to engage in deliberation about the nature of the (perceived) offense. If Todd is not totally oblivious to the nature of the situation, he has the evidential material necessary to reason himself to the belief that there was no, or only a minor, offense. Given that his anger is still responsive to reasons, this belief will have influence on the intensity of his anger, and eventually lead to its fading. Yet it is important to see that, *at the point when Todd becomes angry*, he

could not have avoided his anger by any reasonable means available to him, and seems thus to be fully excused.

Why does it seem as if Todd is fully excused? My explanation is that he is, in fact, blameless_{SH}. To see this, consider first a plausible line of thought that tempts us to judge that Todd is blameless (period). As I noted above, Todd seems to be in a dilemma-like situation. However, it is not that, no matter how he would act, he acts wrongly. Rather, by performing his duty to take care of his mother, he does what he *ought* to do. It seems that you cannot end up blameworthy if all you do is what you ought to do. From this, it follows that one is not necessarily blameworthy for having attitudes that are irrational, even if they are harmful to others. For one might end up with such attitudes even though one always does what one ought to do.

To illustrate this line of argument more clearly, consider the principle of the *transmission of permission*:

(TP) If it is ok for you not to avoid an attitude, or to cause yourself to have an attitude, or to maintain an attitude, then it is also ok for you to have this attitude (for the moment).¹⁵

The principle gains its plausibility, once more, from the idea that having an attitude is a kind of state we are in.¹⁶ For the principle is obviously true if we put ‘ordinary state,’ like being outside, in the place of ‘attitude.’ Imagine a person with authority over your whereabouts is telling you ‘It is ok if you go outside.’ Feeling confident, you end up outside later. Then this person tells you ‘Why the hell are you outside?! I did not tell you that you are allowed to *be* outside!’ We would surely feel like this person is messing with us.¹⁷

We could imagine cases in which a slightly different accusation could make sense. Imagine that you were told that you are allowed to go outside. You go outside and linger around there for very long. Assume that, for some reason, lingering around there for a longer period is offensive to someone. You could later be accused for lingering around, although you were allowed to bring you into the state of being outside. The addition in brackets at the end of (TP) is meant to rule out such cases. Todd might well be blamed for not managing his angry response once it took place. But it seems that he is not blameworthy for it taking place.

¹⁵ Note that (TP) is a variant of the principle of Deontic Action-Attitude Transmission (AAT) which I discussed at length in chapter 6.

¹⁶ As I have argued in ch. 9.1.

¹⁷ Cf. ch. 6.2.1.

Thus, it seems that permissions to not avoid being in a state, or causing oneself to be in a state, transmit to being in a state itself. As argued in chapter 6, the principle that requirements to cause oneself to have an attitude create permissions to have the attitude is true insofar as the permission to have the attitude is not understood as a permission of rationality. For the attitude we are required to cause can still be irrational. Such irrational attitudes can have significant consequences for our interpersonal relationships. However, the sense in which it is impermissible to have the attitude is that you violated a standard of rationality: in this sense, it is ‘impermissible’ to believe against your evidence even if it saves the world.

However, if we do not talk about the standards of rationality, but rather about indirect requirements to have or not to have certain attitudes, then principles like (TP) or (AAT) (from chapter 6) turn out to be true. This is rather trivial, however, since ‘that it is ok for you to have this attitude’ then just *means* that you are not required to exercise indirect control over your mind so as to avoid the attitude.¹⁸

Applying (TP) in this sense to Todd’s case gives us the result that it is, in a sense, ok for Todd to be angry in the face of no, or only a minor, offense: it is ok for Todd not to avoid his anger. More generally, whenever a subject is allowed not to avoid an attitude, or to cause themselves to have an attitude, the subject is also, in a trivial sense, allowed to have the attitude: the attitude is just part of the state the subject was allowed to cause, and it is not appropriate to, in a sense, blame people for states they were allowed to cause. Thus, the implications of this for Todd’s hypological status do not seem to be as trivial: for Todd is blameless (in a sense) for his irrational and harmful attitude.

However, the senses of ‘permission’ and ‘blameworthy’ I employ here are not the only senses of these words relevant here. Susan Wolf distinguishes between a ‘superficial’ form of blame and a ‘deep’ form of blame (1990, 64). In the superficial form, to blame someone just *is* to note that the person is a ‘bad act-maker’ – someone who is not good in acting. In the ‘deep’ form, blame carries a specific ‘force,’ which is naturally interpreted as imposing some kind of harm or punishment on the blamed party. Hieronymi (2004) argues that the latter view of the force of blame gives rise to the idea that it is unfair, and inappropriate, to blame someone with a harsh life. According to her, blame should not be conceived as a kind of voluntary action that would, under certain circumstances, amount to punishment, and thus there is nothing

¹⁸ Cf. ch. 7.1 for the analysis of ordinary statements of the form ‘S ought to have an attitude M’ as ‘S ought to cause themselves to have M or maintain M by exercising indirect control.’

inappropriate about responding to an irrational and harmful attitude with blame, because a blaming emotion is no voluntary action with which we voluntarily impose a burden on another person.¹⁹

It is important to see here that the reason why I regard Todd as being not blameworthy in a sense is not that blaming him would be unfair punishment. Rather, the reason is that, from Todd's perspective, he did everything right, and his permission to act in the way he did transfers to the attitudes he ends up having. There is an important sense in which Todd is not at all a 'bad act-maker' – he might be very good in doing the right things in general. As will become clear in the next chapter, certain *severe emotional responses* towards Todd are inappropriate insofar as one knows about his unfortunate situation.

There *are* ways of appropriately reacting to Todd that could be identified with Wolf's 'superficial blame' – like not trusting in his emotional reactions, treating him with a certain care so that one does not cause him to do harm to anyone, maybe avoiding his company, or even not wanting to be friends with him and thinking badly of him as a person insofar as he is so quick to lose his temper. Some of reactions seem to be close the ones we retreat to when we notice that someone is not responsible – they might be indicators that we adopted Strawson's 'objective attitude,' and thus do not respond with reactive emotions like resentment, anger, or indignation. These responses might indeed be viewed as 'superficial' and not essential to blame. Yet as I will argue in the next chapter, there are some reactions appropriate to individuals like Todd that we only take towards responsible beings. I think distrust, viewing a relationship as impaired, and thinking badly of someone as a person are some of these responses.

To clarify once more, my argument for the claim that Todd is not blameworthy in one sense of 'blameworthy' should be understood as follows: there is one way in which (TP) is obviously true. People are, in a sense, blameless for a state if they were allowed to cause that state. It does not matter for their blameworthiness, in this sense, whether their state of mind is part of this state that resulted from their actions. My argument is just meant to establish that this sense of 'blameworthiness' exists. I will have more to say about this sense of blameworthiness in the next chapter. A clearer grip on it will allow us to see, first, in what way individuals like Todd – and I will extend Todd's case to structurally similar cases which involve victim criminals, psychopaths, racists, or sexists – are not blameworthy, and, secondly, in which sense

¹⁹ Though it is important to see that we can instrumentalize our emotions in order to punish someone – if we do so, then our emotional response will usually take a slightly different form which will be noticeable by the person who we intend to punish with our emotion. Thanks to Pamela Hieronymi for pointing this out to me.

they could still be blameworthy, and, thirdly, in what way we ought to respond to them with our own attitudes and actions.

10.6 SUMMARY

This chapter has brought us back to the concept of responsibility and blameworthiness with which the original puzzle is concerned (10.1). I have argued that certain accounts of responsibility for attitudes imply that we are responsible for our attitudes in two ways that are irreducible to one another: we are directly answerable and historical responsibility for them (10.2-10.3). One is directly answerable for one's attitude is, in a minimal sense, responsive to reasons: it is something for which it is appropriate for others to ask us for the object-given reasons for this attitude. One is historically responsible insofar as one could have managed one's attitude through one's actions. Furthermore, I have argued that if the two kinds of responsibility are indeed as distinct as I argue, then we might have some reason to doubt that answerability is a genuine form of responsibility (10.4). However, I suggested that we can solve this problem by acknowledging an ambiguity in our concepts of 'ought' and 'reason' that corresponds to the two kinds of responsibility at issue. Finally, I have argued that acknowledging historical responsibility as distinct from answerability can explain our intuition that individuals like Todd or Ronja are blameless for their attitude, even though, in fact, they are not fully blameless: they are not blameless_A, but only blameless in the historical sense (10.5). I have thereby provided some characterization of historical blameworthiness and defended the philosophical utility of this concept. Let us now have a closer look at the two concepts of responsibility, and how we might make it plausible that answerability is in fact a form of responsibility (as called into question by the argument in 10.4).

Chapter 11***Two Faces of Responsibility for Attitudes***

In this chapter, I will clarify how I deny premise (3) of the argument against direct responsibility for attitudes presented in chapter 10.4. The premise states: For any attitude *M*, it cannot both be true that we ought not to have *M* but are allowed to have *M*.

I argue against (3) by defending both indirect responsibility and answerability as genuine forms of responsibility that are associated with two distinct kinds of reasons. The forms of responsibility are *genuine* insofar as they are systematically connected to different sets of reactive attitudes. We can be blameworthy both for our violations of requirements of voluntary conduct as well as our violations of the standards of rationality – i.e., of failing to respond correctly to our object-given reasons. Yet both kinds of failures are different in nature, and thus presuppose different kinds of responsibility and give rise to corresponding different faces of blameworthiness. Thus, both kinds of failures violate different kinds of ‘ought,’ and we are able to conceive of cases where we are, in a sense, required not to have an attitude and yet, in a sense, allowed to have it.

I begin with my central argument that we are directly responsible for our attitudes in virtue of being directly answerable for them: the argument from apology (11.1). I then turn to my diagnosis that there are two concepts of responsibility and blame (11.2): Introducing these two senses will cast light into the dispute philosophers take themselves to have when discussing about the nature of responsibility and blame: often, both parties just talk past each other. I suggest instead that they should talk *with* each other, and thus incorporate the insights of the other party within their overall accounts of responsibility and blame. I support my diagnosis by presenting cases where it is plausible that our passionate blaming-responses rationally diminish insofar as we acknowledge a person’s lack of indirect control over an attitude (11.3). I then discuss the two most common objections against this claim – one coming from within mental involuntarism (i.e., the position I defend), and one coming from indirect mental voluntarism (11.4). I provide a further argument for my diagnosis by showing how it might help us to understand traditional puzzles of responsibility: victim criminals and cases of moral luck (11.5). I finally summarize my solution to the problem of mental responsibility (11.6).

11.1 THE ARGUMENT FROM APOLOGY

Denying that answerability is a genuine form of direct responsibility for attitudes conflicts with our practice of blaming people for disrespectful attitudes, like racism or sexism. Blaming the disrespectful seems to be appropriate independently of whether they could manage their disrespect. As Basu/Schroeder (2019) point out, our blaming responses to disrespectful attitudes like racism and sexism seem to be directed towards people *for having these attitudes* rather than for their failure to *manage* their mental lives.

We can confirm Basu/Schroeder's observation by considering the appropriateness of *apologies*. Even if a racist or sexist had no opportunity to manage their attitude, they owe an apology to the disrespected upon realizing their disrespect. Thus, in a sense, the disrespectful are blameworthy, and thus responsible, merely in virtue of their disrespect. Put in a syllogism:

- (1) If S disrespects P in a way noticeable to P, then S owes P an apology.
- (2) If S owes P an apology, then S is blameworthy.
- (3) If S is blameworthy, then S is responsible.
- (4) Thus: If S disrespects P in a way noticeable to P, then S is responsible.

'Disrespecting P in a way noticeable to P' here means that one's disrespect manifests itself involuntarily (think of such involuntary reactions as eye-rolling). It is important for the argument that it does not mean that one performs a disrespectful *action*, that is, it does not mean that one exercises voluntary control over the behavior that is explained by one's disrespectful attitude. Otherwise the argument would not support that answerability is a form of responsibility. Understood properly, however, the conclusion states that merely *having* a disrespectful attitude (that is noticeable) implies that one is responsible.¹ This contradicts *The easy solution* as well as anti-responsibilism. It establishes that S is *directly* responsible for their attitude. Given the further premise that the direct responsibility at play in the argument is answerability, it follows that answerability is a form of responsibility:

Direct Responsibility is Answerability: Being directly responsible for one's attitude is to be answerable for it, i.e., to be liable to inquiries for one's reasons for the attitude.

Why is it plausible to interpret the direct responsibility at play in the argument from apology as answerability? This is because the disrespectful person is answerable for their disrespectful

¹ A disrespectful attitude can be an offensive belief, a malicious desire, a hostile emotion, or an evil intention, for example.

attitude as well as the disrespectful behavior that results from it. We can ask them: ‘Why did you roll your eyes upon P’s remark?’ We can do so even though S’s eye-rolling was not a voluntary action, but rather the result of S’s disrespectful attitude. If someone is involuntarily rolling their eyes on one of our remarks, the person is liable to our inquiries for their reasons for doing so, and if it turns out that they did so because they do not take us seriously (rather than, say, because a fly irritated their eyes), then they would ordinarily be incapable of justifying their behavior and the underlying attitude towards us.² The person’s blameworthiness will be determined by the answer they would provide for their behavior. Thus, the responsibility at issue here is answerability.³

But isn’t this unfair? Why *condemn* someone who did not have an opportunity to avoid becoming who they are, and thus ending up with a disrespectful attitude that sometimes results in disrespectful behavior? However, I have already noted at the end of chapter 10.5, following Hieronymi (2004), that the charge of the unfairness of our blaming responses only comes up if we conceive of our blame as *punishment* – in that way, my talk about blame as ‘condemnation’ might be misleading. To punish someone means to voluntarily impose a burden on the person. But neither do we always impose burdens with our blaming responses (as will become clear when I discuss these responses in more detail below), nor do we always voluntarily instrumentalize our blaming-emotions in order to impose a burden on others. Rather, our blaming-responses are often involuntary and rational reactions towards someone else’s failure of proper regard, and thus cannot be conceived of as fair or unfair.

Let us consider the premises of the argument from apology in turn. I will take (3) for granted: it just expresses a basic assumption about the conceptual relations between blameworthiness and responsibility as the term is employed by most philosophers.

To see why (1) is plausible, consider how it would be inappropriate for the S to, instead apologizing to P, trying to excuse themselves by saying ‘I was just so stressed out today that I

² In the fly-case mentioned in the brackets, the person might turn out that the person was not even liable to inquiries for their reasons, because it might have been a mere reflex – i.e., no involuntary behavior that reflects an attitude, and which is thus completely outside the space of reasons.

³ There might be some dispute about whether an attitude that is disrespectful is necessarily irrational. However, we need not decide this question here. As I have argued in 10.3, it is unproblematic for me to focus on cases in which the attitude is irrational. This will get also those philosophers on board who argue that we cannot be blameworthy if we are fully rational (a group to which I tend to count myself to). If we could be blameworthy for fully rational behavior, then this would imply that we would appropriately judge the person blameworthy if their answer to our request for reasons would be unsatisfying for *in our lights*. By contrast, if we deny that fully rational behavior can be blameworthy, then the appropriateness of our hypothetical judgments will be a matter of the reasons *they* have. In both cases, we could save the idea that direct answerability is a form of responsibility – it is just that the answerability would be either to the reasons the person who judges or to the reasons the judged person has.

could not manage my attitude towards you.’ This response would be beside the point: it is a self-centered response that cares more about one’s own blamelessness than about the other’s standing (cf. Hieronymi 2019). Looking for excuses in this way is not an appropriate reaction towards a person you disrespected. Rather, the disrespectful owe the disrespected an unconditional apology. A sincere apology can repair the relationship that was impaired by their disrespectful attitude. That the disrespectful owe this apology implies, in a sense, that they are blameworthy, and thus responsible. The blameworthiness at issue is blameworthiness_A, and the responsibility is answerability.

One could object against (1) that excusing oneself in the way I called ‘self-centered’ is not inappropriate. Rather, if it is literally true that S could not have managed their attitude, then S is blameless for their attitude, and thus does not owe an apology to P, but rather an explanation of their behavior by providing an excuse.

However, attending to the concept of blame_A – the marking of impaired relationships that is the essence of answerability-blame – allows for an explanation of why one owes an apology rather than an excuse: only a sincere apology can repair the relationship that was impaired by disrespect. If one merely tries to excuse oneself, one does not acknowledge that there is an impairment. Yet disrespect creates such impairments, no matter how the disrespect came about. Take *Ronja the Racist* from chapter 10. Tragically, it would not help her relationship towards the disrespected if Ronja merely points out: ‘But I had no choice!’ This might be the start of a conversation with the disrespected (if they are patient), but it cannot be the end.

We could argue against (2) by granting that S owes P an apology, but by trying to explain this without committing to S’s blameworthiness. Maybe the reason why it is beside the point for S to merely excuse themselves rather than to apologize is not that they are blameworthy, but rather that we have a social practice where we expect each other to apologize even for things we are not genuinely blameworthy for.

Proponents of this objection would have to argue that it is appropriate for Ronja to apologize to the disrespected even though she is not blameworthy for what she should apologize for. The only sense of ‘appropriate’ they can spell out here seems to be pragmatic: within our current society, we contingently expect one another to apologize for disrespect. But we can ask further: why is that expectation justified? The objector against (2) would have to give an account of potential benefits of the existence of such a practice in order to give meaning to the idea that it is ‘appropriate’ to apologize.

However, even if we can give such a pragmatic interpretation of ‘appropriate,’ it would not be the sense in which it is appropriate to apologize for disrespect. We can have pragmatic

reasons *not* to apologize for our disrespect (if we do not want to exclude ourselves from a group of racists in which we grew up, for example). But not apologizing would still not be appropriate in the relevant sense of ‘appropriate’: there in fact *is* an impaired relationship with the disrespected, and this fact constitutively calls for an apology even if there are pragmatic reasons not to apologize.⁴ Thus, the second reply fails because it can only explain the appropriateness of apologies by changing the meaning of ‘appropriate’ – and thereby changing the topic.

I thus conclude that the argument from apology is sound, and that we are directly responsible for our attitudes in virtue of being directly answerable for them. Direct answerability is a form of responsibility, and if we cannot provide adequate object-given reasons for our attitudes that would justify them (at least from *our* perspective),⁵ we are often blameworthy (in a sense). This sense of blameworthiness gives meaning to the idea that we always ‘ought’ to respond correctly to our reasons for attitudes: we can be blameworthy merely in virtue of the fact that we fail to respond correctly to these reasons, e.g., by being (irrationally) disrespectful. Thus, there must be a normative standard in place that says that we *ought* to respond correctly to our object-given reasons.

A legitimate worry with my argument is that it focusses on respect. One might point out that, even if the argument is sound, it merely establishes the normative force of reasons for respect, but not the normative force of reasons for other attitudes. We might conceive of a position that grants that we can be directly blameworthy for some attitudes, but not for others. According to this position, there might be, for example, no epistemic blame: we would never be directly blameworthy for our beliefs merely in virtue of our failure to properly base our belief on our evidence. Rather, we would be blameworthy only if our belief manifests disrespect or if we failed to properly manage our doxastic life.

My main worry with this position is that it is odd. It grants that object-given reasons for attitudes have normative force (and can thus legitimately be considered as ‘reasons’) when we talk about the attitude of respect. But it denies the normative force of object-given reasons (and thus ultimately their status as *reasons*) when we talk about other attitudes. However, there seem to be too many structural analogies between reasons for attitudes. The reasons in virtue of which we are blameworthy for disrespect are object-given: they are provided by the constitutive

⁴ How can these constitutive reasons be weighed against what I called ‘pragmatic reasons’ in order to determine whether we should apologize, all things considered? My suggestion is that these constitutive reasons are, in this case, of the same kind as the pragmatic reasons: they are provided by value (the value of the relationship and the general value of promoting a society where we respect one another).

⁵ On my position on the perspective-dependency of blameworthiness, cf. my discussion of *Ronja the Racist* in ch. 10.3.

aims of attitudes, and failure to comply with them can make us irrational. The same holds for object-given reasons for belief, desire, emotion, and intention. Because of these structural analogies, I deem it to be more promising to explore how we can be blameworthy for violations of epistemic norms and of other norms of rationality, rather than accepting a scattered account according to which we can be blameworthy only for some failures of rationality, but not for others.

Furthermore, it is doubtful that respect and disrespect make up an own category of attitudes for which we can have reasons.⁶ Being disrespectful seems to consist in having specific beliefs, desires, emotions, and intentions by which we sometimes wrong others when these attitudes manifest themselves in our behavior, whether voluntary or not. A person's belief that you belong to a group that should not be taken seriously in political conversations can be an instance of disrespect. On the plausible assumption that when one is blameworthy, one is blameworthy in virtue of the fact that one fails to respond correctly to one's reasons (what else should blame be grounded in?), it follows that the person is blameworthy for their disrespectful belief in virtue of their irrationality: the person can be blamed for their belief because it is epistemically irrational. The same argument can be made for any other attitude that can be disrespectful.

I thus propose that we should understand the faces of blameworthiness better in order to see the nature of direct blameworthiness for attitudes, and how it differs from indirect blameworthiness for attitudes.

11.2 THE DIAGNOSIS: TWO FACES OF BLAME

The difference between the two faces of blameworthiness I want to draw our attention to here can be spelled out by focusing on the kinds of attitudinal and emotional changes the kinds of blameworthiness make appropriate: while the involuntarist face of blame (blame_A) is essentially linked to changes in the kind of relationship we have towards the blamed party, the indirect voluntarist face of blame (blame_H) is essentially linked to the passionate reactions – the 'moral sentiments' (Wallace 1994) – which became the focus of philosophical discussion about responsibility since Strawson (1962).⁷ Distinguishing these two faces of blame will, first, clarify in which sense we are not, and in which sense we are, directly responsible for our mental

⁶ I am grateful to Matthew Chrisman for pointing this out to me.

⁷ Strawson also focused on the role of blame in relationships to understand the nature of blame. The accounts of blame I discuss in what follows can both be viewed as broadly Strawsonian.

states; it will secondly cast light into both the seeming dispute between the nature of blame as well as the seeming dispute between defenders and deniers of direct responsibility for attitudes; and it will thirdly provide the material to resolve paradoxes of responsibility, namely, first, what attitudes we should have towards psychopaths, victim criminals, racists and sexist which became so as a result of structural forces within society, as well as, secondly, our ambiguous reaction to moral luck cases.

While these remarks of what this chapter is meant to achieve sound very ambitious, I think that the points I make are not too controversial given the recent philosophical discussion on blame. I will start in this subchapter with the second point I made (on seeming disputes), and turn to the other issues in the subsequent subchapters.

Consider two opposing parties in the discussion about the nature of blame. The first party, to which I count philosophers like Hieronymi, Scanlon, and Smith, emphasize how blame's primary function is to mark the impairment of relationships. I view them as defending blame as being essentially the concept of answerability-blame (blame_A). Given this face of blame, it is hard to see why control or reasonable opportunities should be necessary for the blameworthiness of a person: as soon as our relationship with someone is in the relevant ways impaired because of their failure to respond correctly to their reasons, and we notice that impairment, there is blame involved, no matter whether the reasons for this impairment stem from violating requirements of voluntary conduct or involuntary flaws in someone's character. In fact, viewing the relationship as impaired *just is* the blame. Since we are obviously justified in treating our (moral) relationship with, e.g., victim criminals or *Ronja the Racist* (cf. chapter 10.2) as impaired in the relevant sense, there is also a sense in which we are justified to blame these people.

The second party, to which I count philosophers like Fischer,⁸ Wallace, or Watson,⁹ view blame as having a more serious 'opprobrium' (Wallace 2011, 349) or 'force' than accounts of the first kind allow for: to blame someone is not merely to view a relationship as impaired, but rather to view it as being impaired in a specific way that calls for specific reactive emotions. According to Wallace, 'it would indeed be strange to suppose that one might blame another person without feeling an attitude of indignation or resentment toward the person, or that one might blame oneself without feeling guilt' (1994, 75). According to this account, these

⁸ Including his co-authors Ravizza (Fischer/Ravizza 1998) and Tognazzini (Fischer/Tognazzini 2009).

⁹ I include Watson insofar as he talks about accountability rather than mere attributability (Watson 1996). In fact, what I say in what follows is close to Watson's distinction between these two forms of responsibility. It can be viewed as a sketchy attempt to spell out the difference between these two forms.

more *passionate* emotions are essential to blame. My diagnosis involves the idea that these philosophers, insofar as their theories are meant to be also about the blameworthiness of attitudes, talk about blame_H.

As Smith points out (2013, 32), philosophers of the first party object to the view of the second party that blame does not necessarily involve these emotional components. Especially in the case of blaming loved ones for moral failures, we might do so without hostile emotions. Furthermore, we might note a modification in our moral relationship to someone by ‘dispassionately “unfriending” someone on one’s Facebook page, for example, or by simply refusing to trust anymore, and these too should qualify as blame’ (ibid.). However, I think Wallace’s point about the strangeness of counting these responses, insofar as they are done dispassionately, as forms of blame, has some validity. My analysis here will imply that the passionate element is only appropriate when people who have wronged others in their voluntary conduct – more precisely, by failing to make an effort of will they owe to others. The passionate element is thus not made appropriate *merely* by having an irrational attitude that impairs a relationship. This will do justice to both Wallace’s point as well as to accounts defended by philosophers of the second party, who do not see the passionate element as necessary for blame.

As I made clear in subchapter 10.5, there is a sense of ‘blameworthy’ (the sense of historical blameworthiness, blameworthiness_H), in which people whose attitudes came about in such a way that these people had no reasonable opportunity to avoid having these attitudes *cease* to be blameworthy. My suggestion is that proponents of the second party have this sense of blameworthiness in mind when they defend their account of blame. Certain kinds of emotional reaction, they point out correctly (e.g., Fischer/Tognazzini 2009), will not be appropriate towards people who suffered a certain kind of history, even if they hold irrational attitudes that impair relationships as a result of this history. For example, it is not appropriate to *resent* or to be *angry* with or to be *disappointed* of Ronja, given that one knows about her history.¹⁰ Nevertheless, it would still be appropriate, if one was Ronja’s friend (for example), to modify one’s relationship towards her when one finds out that she holds racist attitudes.

¹⁰ Though anger raises a set of difficult questions: If one is directly affected by her racism because one is its target, would it not be justified to be angry at her just for *being* racist? There are at least two things I can say in reply. First, according to the account of indirect mental requirements defended in chapter 7.1, there *is* a sense in which the targeted person is allowed or could, in principle, even be required to feel anger: being angry might make others aware of the injustice, or it might promote the social cause to keep racism within its limits. These reasons would justify anger in the sense of justifying any actions that promote this kind of anger. Secondly, it seems that even the targeted person should be affected in their anger insofar as they are aware of Ronja’s history and insofar as they are rational. The fitting response they should give to someone who has a less involuntary history than Ronja would be more emotionally severe.

If my proposed diagnoses here is right, then this has implications. First, since there are two different *concepts* of blame involved, the two parties mentioned in the last two paragraphs just *talk past each other*.¹¹ Our word ‘blame’ does neither have *one* static use in ordinary language nor a very frequent use (at least nowadays). The metaphilosophical implications of this fact are not properly taken into account in this discussion. For if a word does not obviously correspond to just one concept and is not used frequently, then it is far from clear that we can just do *conceptual analysis* in order to clarify the nature of what the word refers to. To do conceptual analysis in this sense means to identify the conditions of proper application of this word given how we use it in ordinary language. We do so by relying on intuitions of competent speakers who can correctly apply the word. However, if a word might have multiple meanings and we barely use it, then we might have intuitions that are misleading (due to the ambiguity of the word) as well as not much intuitions on which to base our judgments of proper application (due to its rare usage). We should then not be surprised if different accounts of ‘the nature’ of blame come up as a result of our method of conceptual analysis. Rather, this is to be expected if the use of the word is complex and can probably not be captured by just one of the resulting analyses.

I thus propose that what the discussion about the nature of blame is about cannot be our ordinary concept of blame, and thus the proponents cannot aim at a conceptual analysis – talk of ‘blame’ is, at least nowadays, a philosophical artefact. What we can instead do, however, is to analyze the different kinds of responses that philosophers summarize under the label ‘blame’ (or ‘criticism’). This enterprise, however, might result in very nuanced accounts of different kinds of responses which have a kind of family resemblance, but no unifying properties. And the conditions under which these responses are appropriate might differ significantly.

So the discussion on blame should, I suggest, instead be conceived as *conceptual explication*: we seek to identify a *useful* use of the word ‘blame,’ given our theoretical (or practical) purposes. And of course, different concepts of blame might be useful for different purposes. Though our starting point is restricted by certain intuitions about the concept (e.g. that it is a form of personal criticism that is neither mere evaluation nor a form of punishment, see Scanlon 2008, 122), how we use the concept is, ultimately, up to us. Given that this is what we do when we discuss about ‘the nature’ of blame,¹² distinguishing between two concepts of

¹¹ This is, of course, not to say that their discussion contains very valuable insights and philosophical progress. Rather, it just means that the way the discussion is framed is misconceived.

¹² I put ‘the nature’ in quotation marks because, given the kind of enterprise conceptual explication is, it is unclear what ‘nature’-talk is supposed to be about – the nature of anything (at least anything that is not a natural kind) can only be ‘a shadow cast by our grammar’ (Hacker 1990, 438): necessary conditions on proper application

blame so as to spell out in what sense a person can be (directly) blameworthy for their attitudes and in which sense a person cannot be (directly) blameworthy for them is completely innocuous. I thus conclude that, if philosophers disagree about blame's 'nature' without having in mind a specific common purpose for which to use this concept, then their disagreement is merely verbal.

11.3 SUPPORTING THE DIAGNOSIS: THE RATIONALITY OF PASSIONATE BLAME

Does this mean that there is no substantial disagreement (but only verbal disagreement) philosophers like Hieronymi, Owens, and Smith, who view blame as the marker of impaired relationships, and philosophers, like Fischer, Wallace, and Watson, who view blame as essentially involving passionate emotions? In some sense, it does. Insofar as Hieronymi, Owens, and Smith merely talk about impaired relationships when they talk about blaming people for their beliefs, emotions, and other attitudes, it is undisputable that our relationships towards racists like Ronja can be legitimately impaired because of her attitudes, and that we can thus legitimately *blame* Ronja in precisely this sense: we might no longer view her as a close friend, we might distrust her in certain contexts, we might be justified in not inviting her to certain events, and we might cease to promote certain of her projects or no longer take pleasure in her successes in the same way we did before. If this is *all* these philosophers want to say about blameworthiness for attitudes, there is nothing controversial about the claim that people can be blameworthy merely for *having* certain attitudes. We need not even decide the question of whether having an attitude essentially involves some kind of non-voluntary agency in order to understand that people *are* sometimes directly blameworthy for their attitudes in this sense: their attitudes make up or reveal something about their character, and what they make up or reveal gives us in turn reasons for various responses to other people.

The appropriateness of these reactions marks a difference in possible relationships towards beings like Ronja, or the victim criminal, or psychopaths, on the one hand, and beings like animals, on the other. We do sometimes 'not trust' animals in the sense that they might, e.g., bite us. Genuinely distrusting Ronja or other persons with a problematic history who wrong others has a different 'flair' towards distrusting an animal – we only judge the former's behavior to be rooted in a bad character, but not the latter's. We can thus view our relationship

just derive from how we use a word. However, if our grammar does not give clear conditions of application, then our grammar does not even cast much of such a shadow (beyond some initial intuitions that restrict possible uses of the word).

to Ronja etc. as impaired in virtue of this character, but not our relationship towards non-responsible beings. We do not trust the animal in the sense that their behavior is not fully *reliable*. But we do not trust the persons in the sense that they might not live up to our *expectations* or *demands* which we have towards most human adults – expectations and demand which we have even if they are generally *unreliable*. We might well know about Ronja's racism and how she will behave as a consequence of her attitude, thereby being able to reliably predict what she will do. Yet we could still be said to distrust her.

What I deny is that it is fitting to *resent* Ronja merely for her beliefs, or to be severely *angry* at her while knowing her involuntary history.¹³ I would argue that it is not even appropriate to be *disappointed* of her among finding out that she has racist attitudes as a result of this history. Why be disappointed of Ronja? What should she have done otherwise? As Scanlon notes in a footnote, there is a sense in which people like her are not blameworthy insofar as it is *not the fault* of them that they ended up being the way they are (2018, 93, fn. 9).¹⁴ Even if they end up being irrational and vicious, this might not have been their fault in an important sense. Being disappointed might be appropriate if Ronja tried to intentionally *deceive* us about her attitudes – if she, upon reflection, was aware of her racism but made us believe that she did not have racist attitudes. What we are appropriately disappointed about then are her intentional attempts to deceive us about her attitudes, not her attitudes themselves. Insofar as defenders of direct responsibility for attitudes want to say that *these* kinds of emotional reactions are appropriate reactions towards Ronja's racism, I disagree with them.

Notice how it would be too harsh of a self-judgment if Todd was *disappointed* with himself after becoming aware of his inappropriate anger. If Todd was fully self-aware, he might reasonably expect that he will end up in a situation in which he feels inappropriate anger. Although it is then appropriate for him to sincerely *apologize* for his angry reaction (which is a reflection of the fact that, *in some sense*, he *is* blameworthy for it: there is an impaired relationship that needs to be repaired), it would be *a further step* for him to feel disappointment or even guilt for his reaction.¹⁵ This further step is only justified if he could have reasonably avoided his anger. Otherwise his attitudes towards himself would involve judgmentalism. Similarly,

¹³ On anger, see fn. 10 above.

¹⁴ To be precise, he claims that people who cannot, because of their beliefs or incapability to take in a certain perspective (like the moral perspective), see that they are acting wrongly, are blameless in the sense that it is not their fault that they cannot see this. See my discussion of *Ronja the Racist* in 10.3: It is unclear to me whether Ronja can be blameworthy if we understand her case in such a way that she lacks any evidence that her views are wrong. Yet there might be a sense in which Ronja, even if she has sufficient evidence, is still *unable* to see where her evidence points to: her irrationality was deeply ingrained by the society she was raised in.

¹⁵ I assume that guilt is the self-directed analogue to resentment, anger, or disappointment.

others who know about Todd's situation might reasonably expect him to become angry at them while they are around him. Insofar as they know about his situation, I want to suggest that it is as irrational to feel indignation or resentment towards Todd in the same way as it is irrational for Todd to be disappointed with himself or feel guilty about his anger.

What if Ronja, upon reflection, becomes aware of her racist attitudes as something problematic and feels guilty? Wouldn't this be an appropriate reaction to her own attitudes? Although we could not, of course, rule out that she might actually feel guilt for her attitudes, we should not view this emotion as a *fitting* reaction.¹⁶ Although she might *actually* feel guilty for her attitudes, we should not view this emotion as a *rational* reaction. For it is rational for her guilt to diminish to the degree to which she acknowledges her own 'facticity' – i.e., her own subjection to influences beyond her control. Suppose Ronja realizes how she disrespected people in the past and feels guilty about it. She then comes to understand her own background better, and how her disrespectful attitudes came about through no fault of her own. It would be very natural for her guilt to diminish as her understanding of her facticity increases. If we were to deny the rationality of her modified feeling of guilt, we would have to view this process of gradual diminishment as not rational. But this does not seem plausible. Acknowledging the fact that she had no such opportunities should rationally lead Ronja to quite different reactions than guilt – like distrusting some of her own judgments, and thus desiring to become a different person.

Consider the following objection to my view that the passionate form of blame is not a rational reaction towards wholly involuntary attitudes. Ronja's attitude is a threat to the *standing* of others – even though it is as involuntary as it can be, and probably it is a threat especially *because* of its involuntariness. So what if a person who was the target of Ronja's attitude has to encounter her every day and suffer from the way she is treating them? Surely, this person has some standing to blame her, and, so the objection goes, has all the right in the world to feel passionate responses like anger, resentment, and indignation. These emotions might be a way to express their *moral protest* against racist attitudes.¹⁷ However, it seems that this is exactly what I have to deny.

In reply, it is first important to see that I do not deny that anger can be *all-things-considered* allowed or even required as a response to involuntary racism: we should sometimes *ensure* that we show our moral protest, even if the racist offense is a result of an involuntary

¹⁶ Though it might be *justified* in an all-things-considered sense – see the fn. 10 on anger above.

¹⁷ On blame as a form of moral protest, cf. Smith (2013).

history.¹⁸ In this sense, I can just agree that the victim has a *right* to intense passionate responses. As Ceshire Calhoun points out,

our interest in questions of moral responsibility is more than an intellectual one, satisfiable by achieving correct judgments about responsibility. We also take a practical interest in determining how we ought to respond to wrongdoers. [...] [I]n abnormal moral contexts our entitlement to respond with moral reproach is independent of the blameworthiness of individuals (1989, 400).¹⁹

I agree with Calhoun that we should not view responses like moral reproach as *rational* insofar as these emotional responses are merely directed at wholly involuntary features: in one sense, individuals like Ronja are not blameworthy. Yet in other contexts, it might be all-things-considered reasonable to blame even the wholly blameless, as Calhoun argues.²⁰

A *prima facie* assumption when we encounter people who wrong us is that they could have avoided the wronging (whether they wrong us with their attitudes or their actions). This explains why we are usually *prima facie* justified to react to racist attitudes with passionate emotions. Yet, as I have pointed out already, our natural emotional responses towards these people tend to become less passionate when we realize and see clearly that they could not have avoided their wrong. At the same time, even if we withdraw the passion from our emotions in response to someone's involuntary history, the objective stance would not always be the rational stance. This is because Ronja is, in a sense, blameworthy: the relationship of the wronged party to her is impaired in important ways.

Here is the summary of my argument for the idea that emotions like resentment, indignation, disappointment, (severe) anger, or guilt for an attitude are rational only if the target of these blaming-emotions had reasonable opportunity to avoid this attitude. I have distinguished

¹⁸ Cf. fn. 10 above.

¹⁹ With 'abnormal moral context,' Calhoun refers to the kind of context in which moral knowledge that is commonly shared by everyone is lacking. In such a context, 'the rightness or wrongness of some courses of action [...] are, for a time, transparent only to the knowledge-acquiring subgroup but "opaque" to outsiders. Because moral knowledge is not shared, the presumption that all agents are equally capable of self-legislation breaks down' (1989, 396). Arguably, the case of *ordinary* racism differs from an abnormal moral context because, on a global perspective, there is a growing awareness of racism. Thus, ordinary racists can rationally be blamed passionately because they had plenty of reasonable opportunity to reconsider their attitude. Yet Ronja grew up in a racist community: within this community, there was no awareness of racism, and thus the wrong of racist behavior might have been opaque to the group she was raised in. As I pointed out in ch. 10.3 we must assume here that Ronja has at least the moral knowledge that is necessary for her racism to count as irrational. Otherwise it will not be obvious that she is blameworthy in any sense. Thus, I tend to agree with Calhoun that in a fully opaque context, Ronja might be blameless in every sense of the word, and yet it might be justified all-things-considered to blame her nevertheless.

²⁰ It is important to note that Calhoun seems to think that this sense is the only sense of 'blameworthy' – she thus uses the term in the way proponents of the first party in the debate about the nature of blame do, i.e., as referring to blame's more passionate face.

throughout this and the last chapter between two forms of blameworthiness for attitudes – the blameworthiness of answerability which arises from our failure to respond correctly to reasons for attitudes, and historical blameworthiness which arises from our wronging others by failing to properly make use the opportunities we have for shaping our mental lives. Since, as I have argued at length throughout chapter 10, both kinds of blameworthiness are distinct (comparing them cannot issue in overall judgments of whether a person is blameworthy (period)), it is plausible to suppose that they are associated with two quite different sets of appropriate reactions towards people who are blameworthy in these senses.

My claim that the two different sets of reactions *are* the kinds of sets I suggest – namely, attitudes that mark the impairment of a relationship, on the one hand, and passionate emotions like resentment, guilt, anger, or disappointment, on the other – is inspired by the two different theory-families about ‘the nature’ of blame I sketched above. While the first set of theories plausibly implies that there is direct blameworthiness for attitudes which is independent of any reasonable opportunity to exercise indirect control over one’s attitudes, the second set of theories accepts that indirect control is necessary for responsibility. My suggestion is that the reactions the first theory-family regards as essential to blame map on the reactions that are appropriate towards violating standards of rationality, while the reactions the second theory family regards as essential to blame map on the reactions that are appropriate towards persons who violate requirements not to wrong another person with their voluntary conduct.

My reason for this suggestion is that the second theory-family tends to regard blame as rational or appropriate if there was a violation of a requirement of action involved (if there were reasonable opportunities for the agent to avoid the wrongdoing or the irrational attitude), while the first theory-family does not regard this condition as necessary for blameworthiness. Assuming that proponents of both theory-families capture important aspects of our relationships we have towards other responsible beings, this seems to be a plausible way to combine the insights of both.

Furthermore, my discussions of the cases of Ronja and Todd support the idea that there is an important difference in the way we rationally respond to attitudes that are the result of a wholly involuntary history, and attitudes which a person could have reasonably avoided. The rational responses to attitudes of the first kind, insofar as one is aware of the involuntariness of the attitude, seem to be less passionate than the rational responses to attitudes of the second kind.

What one might be worried about is that I deem *involuntary* responses, like resentment, as appropriate to *voluntary* failures. Instead, one might think that the proper response to

voluntary failures can only be another voluntary activity: namely, *punishment* in order to provide an incentive for the voluntary failure not to happen again.²¹ But what is the point of *resenting* someone for failing to, say, *manage* one's anger?

There are two things to say in response. First, it is not strictly speaking true that I take resentment to be a response to *voluntary* failures. Resentment might still be a response to an attitude a person has towards me. But these attitudes appear in a different light to me when I see the involuntary history of a person. Though this history mitigates the passion of my response insofar as my response is rational, the response can still be directed at the attitude itself – at the racism or the hostile anger, for example, rather than at the failure of the racist or the angry person to manage their attitudes. The person's failure to manage is rather a precondition for my blaming-response that is directed at the attitude to be rational. Secondly, it is not clear to me why involuntary responses to voluntary behavior should never be rational responses to such behavior. For sure, if I realize that someone is just lazy (the paradigm of voluntary failure), being angry at the person might be rational (depending on how often the person is lazy, and on whether the person owes it to me that they make more effort of will).

Below (figure 2), I outline a rough table of the responses I discussed in this chapter, and where to locate them. Both kinds of responses are neither personal evaluation nor punishment or sanction. Neither of these kinds can be said to be conceptually 'closer to' or 'farther away from' either personal evaluation or sanction. They are just different in nature from both. Yet we can distinguish the sets of responses by the passion of emotions that is essentially involved in each kind of response. While what I call 'answerability-blame' below does not require passion, 'historical blame' does. As I have argued, to blame someone in this latter sense (the *passionate* sense of blame) for their attitudes requires that the attitudes were not wholly involuntary. Hieronymi, Scanlon, and Smith were right in pointing out that we sometimes blame people without passion. But they go too far if they want to deny voluntary control as a precondition on this passion.

²¹ Thanks to Pamela Hieronymi for pointing out this objection to me.

mere personal evaluation	answerability-blame change of attitude/intention moral personal evaluation	historical blame passionate emotions / gratitude	punishment
<ul style="list-style-type: none"> - Unreliable - Allowing impairment of relationship (not grounded in character; e.g., no longer playing with a dog who bit you) - Good/bad student or teacher or archer or drinker, etc. 	<ul style="list-style-type: none"> - Distrust (for reasons grounded in a person's character) - Allowing impairment of relationship (for reasons grounded in a person's character), like refusing to make agreements, ceasing to be friends - Giving up intentions to promote their projects - Not taking pleasure in success; not hoping that things go well for the person - Thinking badly of someone as a person - 'stupid,' 'greedy,' etc. 	<ul style="list-style-type: none"> - Resentment - Indignation - Guilt (self-directed) - (Severe) anger - Disappointment 	<ul style="list-style-type: none"> - Prison sentence - "I no longer talk to you"

Figure 2: Forms of personal criticism

11.4 THE RELEVANCE OF REASONS-RESPONSIVENESS AND INDIRECT CONTROL

Although my hybrid account of mental responsibility has the advantage of capturing both of our intuitions about the relevance of responsiveness to reasons *and* indirect control for our responsibility for the attitude, it also comes at a cost. By combining elements of two accounts, I am not merely able to account for more intuitions, I also buy into problems of both accounts. Philosophers that emphasize reasons-responsiveness for an explanation of mental responsibility will complain that I give indirect control too much of a place in such an explanation. Conversely, philosophers who emphasize indirect control will complain that I allow reasons-responsiveness too much space by granting that answerability is a form of direct responsibility. This section will counter these two objections.

11.4.1 IS ONLY RESPONSIVENESS TO REASONS RELEVANT TO MENTAL RESPONSIBILITY?

Owens argues that our responsibility for attitudes is wholly a matter of our responsiveness to reasons.²² His strategy for explaining away our intuition that indirect control is relevant for responsibility and blameworthiness is to say that in cases where a person lacks indirect control

²² Cf. esp. Owens 2000; 2017a, Introduction. More precisely, Owens thinks that reasons-responsiveness is sufficient for our responsibility for beliefs, emotions, and desires, but not for our responsibility for our intentions. In order to be responsible for our intentions, Owens argues, we need to be in 'reflective control' of them. Owens is thus an involuntarist with respect to most attitudes, but a mental compatibilist with respect to intention.

over their attitude, it is usually also the case that their attitude is not fully responsive to reasons. In such cases, the blameworthiness of a person for their attitude is indeed diminished in some way, but this is due to the lack of reasons-responsiveness, rather than to the lack of indirect control. Thus, Owens argues that

[t]he rage of someone terribly abused as a child is less resented than that of a person with a normal upbringing, and temperance in such a person is the more admired. But this should not be taken to indicate that his culpability somehow depends on the degree of control, direct or indirect, which he is thought to exercise over his anger. Rather, what gives us pause are doubts about whether this person's emotions are responsive to reasons at all, given his unusual upbringing. (2000, 120)

The problem with this proposal is that we can imagine cases where a person lacked any reasonable opportunity to exercise indirect control over their attitude, and yet the attitude is fully responsive to reasons. If reasons-responsiveness was all that is relevant to explain why someone is responsible, and to what degree the person is blameworthy for their attitude, then the blameworthiness of the person in such a case would not be diminished by their lack of indirect control. But this is implausible. As I have argued in length above, if someone lacks any reasonable opportunity to manage their irrational attitude, or if they have tried to avoid their irrational attitude by any means possible but failed, then certain passionate responses towards this person cease to be appropriate. The attitude might well be as responsive to reasons as an irrational attitude can possibly be.

Take someone who tries to improve on their irrational angry outbursts. They try very hard indeed, and spend a lot of their free-time in therapy, group-sessions, meditation, and so on. Yet they end up having one of their irrational outbursts again. As Owen's himself notes about a similar case, 'one's reproaches would be tempered by a knowledge of [their] efforts at self-improvement' (118). Yet this can hardly be explained by the fact that the attitude becomes less responsive to reasons in the process of self-improvement. If anything, the attitude becomes *more* responsive to reasons: if the efforts of the person had any effect at all, then there will be more counter-factual scenarios now in which the person would not have had one of their angry outbursts due to their responsiveness to a wider range of reasons against anger.

Since I have argued in chapter 10 that the blaming responses that would be irrational in such a case – resentment, indignation, guilt – presuppose a different *kind* of responsibility (historical responsibility) than the responsibility that is associated with responsiveness to reasons (answerability), I conclude that the responsibility presupposed by the passionate blaming

responses is to be explained in terms of indirect control rather than an attitude's responsiveness to reasons, as Owens claims.

11.4.2 IS ONLY INDIRECT CONTROL RELEVANT TO MENTAL RESPONSIBILITY?

Indirect voluntarists will argue that in cases where an attitude is responsive to reasons, but where the person did not have any reasonable opportunity to manage the attitude, the person is not responsible for it. Meylan (2018, 886-7) presents such a case for the claim that responsiveness to reasons is not sufficient for being responsible:

Sonia has been raised in a sectarian community whose members believe in the statements of some powerful guru called 'Algua.' She, like the other members of her sectarian community, has been taught various very strange things since a young age. Among other things, Algua taught her that dreaming of losing your teeth is an early-warning sign that something bad is going to happen to a member of the sect. Sonia is 17 and the association of these two things is now part of the way she apprehends the world. When she dreams of losing a tooth, she recognizes this to be a reason to believe that something bad is about to happen to a member of the community and believes that something bad is about to happen for this reason. While reading the reliable local newspaper this morning, she remembers having dreamt of some dreadful visit to the dentist that ended with a tooth being extracted. She believes that something bad is going to happen to a member of the community for this reason.

Now, in an alternative scenario in which her reliable newspaper were to contain an article denouncing the absurdity of Algua's teaching, Sonia would recognize this article to be a reason not to believe that something bad is going to happen to her community and she would not believe that something bad is going to happen to her community for this reason. Importantly, the explanation why Sonia would recognize this article to be a reason not to believe that something bad is going to happen to her community is not that Sonia considers newspapers to be creditable sources of information. In fact, the explanation why Sonia would recognize this article to be a reason not to believe that something bad is going to happen to her community is also a matter of her peculiar education. Sonia has been taught since a young age that the more Algua's own credibility is put into question, the better each member of the sect will find herself. This is the explanation why she considers the article to be a reason not to believe that something bad is going to happen to her community.

Although Sonia's beliefs are very different from what most of us believe, we can still conceive of them as being properly based on the epistemic reasons she takes herself to have for them. The 'alternative scenario' is meant to illustrate that Sonia's belief is responsive to reasons in the way required by any reasons-responsiveness account of doxastic responsibility: she would revise her belief in close counter-factual scenarios where the epistemic reasons are different than they are in the actual scenario. The reason why we still regard Sonia not to be responsible

is, according to Meylan, that ‘Sonia [...] is as much the victim of her indoctrination in the alternative as in the actual scenario’ (2018, 888).

I agree that we will not treat Sonia as responsible for her beliefs. The beliefs she infers from her evidence do not count against her as a person: she is not irrational in virtue of drawing these inferences, and she does not, due to the way the example is described, display any epistemic vices like gullibility. Nor could Sonia have reasonably avoided the outlook on the world she has. There is thus no basis for us to blame her.²³ It will be more likely that we would treat her with Strawson’s ‘objective stance’: her ways of making inferences will probably spark our scientific curiosity about the effects of indoctrination, and it might spark philosophical astonishment about the foundation of our own beliefs; furthermore, we might be careful what information we provide her with – but not because we do not *trust* her epistemically (after all, she is quite competent in drawing the right inferences, *given her system of beliefs*), but because we regard her belief-forming processes as *unreliable* in producing true beliefs. Thus, it seems that Meylan’s intuition that Sonia is not responsible is well-grounded.

However, the fact that *we* would treat Sonia this way does not imply that she is not responsible for her beliefs. Note first that people with a radically different world-view from ours would treat us with such an objective stance as well. Members of Sonia’s community will likely seek for explanations of why we believe in the reliability of certain newspapers, and make up conspiracy theories about how *we* are indoctrinated by our surroundings. Our different set of beliefs is not viewed as *challenging* theirs, but rather as an object of investigation. Their objective stance is here a result of the cognitive distance they feel towards us. However, this does not imply that we are not responsible for our beliefs. It just implies that our responsibility-practices are isolated in certain respects from theirs.

Similarly, *within* her community, Sonia’s beliefs will be treated as something for which she is responsible. Depending on whether she is rational in drawing inferences in accordance with Alqua’s teaching, the members of her community will regard themselves as having reason to epistemically trust or distrust her. The relationships towards the members of her community will be affected by how rational or irrational her beliefs are in light of Alqua’s teachings. Since Sonia’s beliefs can have these normative consequences on the relationship with other people, she is responsible for her beliefs in the relevant sense.

That a lack of indirect control over one’s attitudes does not fully undermine our responsibility for them was already established by my argument that answerability is a form of

²³ This holds at least as long as we accept that someone who is fully rational is not blameworthy, cf. ch. 10.3.

responsibility. I have argued in the present section that the reasons-responsiveness of an attitude – i.e., our being answerable for it – is sufficient for being responsible for the attitude even in cases of severe indoctrination. To see the plausibility of this conclusion, we just have to keep in mind the nature of this type of responsibility and remember that it is not a sufficient basis for the passionate blaming responses associated with historical responsibility.

11.5 A FURTHER ARGUMENT FOR THE DIAGNOSIS: MORAL LUCK AND VICTIM CRIMINALS

There is another reason to suppose that the two theory-families about the nature of blame I distinguished throughout chapters 11.2 and 11.3 capture two important aspects of our responsibility practices. If we accept both families as being *complementary* accounts towards an overall theory of responsibility rather than as *competing* accounts, then we have the material to resolve familiar puzzles about blameworthiness which, according to Scanlon (2008, 85), any satisfying account of blame has to solve while also explaining why the puzzles are puzzling to us. Here is Scanlon's description of the two puzzles (about victim criminals, and about moral luck):

[M]ost of us are strongly inclined to think that it is appropriate to blame people who commit terrible crimes. But, on the other hand, we are also sometimes inclined to think that these people cannot properly be blamed if, as seems likely, their characters and actions are caused by factors outside of them, over which they have no control. Similarly, we are inclined to think that two drivers who are equally careless deserve the same amount of blame, but also inclined to think that if one of them kills a pedestrian then he or she should be blamed more severely than the other who, through sheer good luck, gets home without an accident. (Scanlon 2008, 84)

Consider the case of the victim criminal first. It can be appropriate to blame people who commit terrible crimes in the sense that it is rational to view our relationships towards them as impaired, or to sever certain relationships towards those people.²⁴ This explains the first intuition. However, it is not appropriate in *another* sense to blame these people – it not rational to resent them, or to be severely angry at them, merely for their attitudes (and sometimes not even for their actions).²⁵ This is what explains the second intuition about victim criminals which, as it turns

²⁴ Note that, as I discussed in ch. 10.3, the nature of these reactions might differ dramatically whether we assume that the criminal failed to respond correctly to their reasons while forming their attitudes or whether they had such a distorted perspective that they responded correctly to all their reasons in forming their attitudes. In the latter case, answerability-blame might not be appropriate. It is only in the former case that answerability-blame is clearly appropriate.

²⁵ Whether it is rational to resent them for their actions will depend on whether they had the ability to see that their actions are wrong. If they hold a rational belief that their terrible crimes are not wrong, and are unable to

out, only seemingly contradicts the first one: it can be rational to blame them in one sense, but not in another.

Take, next, the cases of moral luck Scanlon describes. The two drivers might well deserve the same amount of blame. Feeling resentment and anger towards them is equally rational in both cases – although, for reasons grounded in human psychology, the degree of resentment and anger of people who are directly affected by reckless driving will sometimes vary, and thus we will, due to the way our mind works, often feel an irrational degree of resentment and anger (see the last footnote above). The second intuition that the driver who actually kills someone due to their reckless driving is more blameworthy is explained by the fact that this driver might be more blameworthy^A. For people closely related to the victim of the driver's recklessness stand in a very different kind of relationship to the driver than they stood to the driver before the accident happened: it is rational to view their relationship towards the driver as impaired in various ways. The request for justification they address to the driver would be a different one than the request the pose to the second one: the first one could be asked 'why did you kill this person?', while the second one can only be asked 'why did you put other's lives at risk?'²⁶

Even for people who are not directly related to the two drivers, other kinds of attitudes will be appropriate towards the driver who killed someone than towards the driver who did not kill someone. The two sets of attitudes, however, will not differ in the degree of resentment or anger. Rather, the attitude towards the driver who killed someone might be *conditional* attitudes such as believing it to be appropriate that the driver takes full responsibility for the consequences of his reckless driving, and a readiness to blame him with further reactions if the driver is not willing to do so.

Thus, by distinguishing between these two concepts of blameworthiness, we can provide initial material for resolving two traditional puzzles about responsibility (I do not claim that I have spelled out a full solution here). This supports the idea that we should distinguish *two concepts* of blame. Since these concepts map on how we blame people for involuntary

see the reasons which speak against their crime (as psychopaths, given certain definitions, are), then it would also be irrational to resent them for their crimes themselves. This does, of course, not imply that people who are directly affected by their crimes are wronging the criminals in feeling resentment towards them. Their feelings of resentment, though not rational, are no wrongs because these emotional reactions are very natural and almost unavoidable. That they *seem* like rational reactions is explained by the fact that people who are directly affected by crimes are not able to see the history of the criminal: their focus is on the terrible crime and the loss they suffered due to it. In the long run, however, we would expect a rational person to adopt another attitude towards the victim criminal insofar as the person understands the specific circumstances of the criminal.

²⁶ Again, in what way it is appropriate to view the relationship as impaired with the driver will depend on whether we think that blame is only appropriate if people failed within their own lights (cf. the footnote before the last one as well as ch. 10.3).

attitudes on the one hand, and how we only blame them for attitudes that were under their indirect control, the distinction is very useful for making my solution to the puzzle of mental responsibility more precise. *The easy solution* is true insofar as we talk about a kind of responsibility that is presupposed by blameworthiness_H. *The easy solution* turns out to be false trivially when we focus on blameworthiness_A – for our relation to others can be impaired in various ways even if they did not violate requirements of action, i.e., without committing a fault in their voluntary conduct, and merely in virtue of their failure to respond correctly to (their) reasons for attitudes.²⁷

11.6 CONCLUSION: THE SOLUTION TO THE PROBLEM OF MENTAL RESPONSIBILITY

Distinguishing between direct responsibility for attitudes as answerability and indirect responsibility for attitudes as historical responsibility (chapter 10) and spelling out the forms of blameworthiness that presuppose the respective forms of responsibility (chapter 11) allows us to see how we can solve the problem of mental responsibility. I will first present my diagnosis of how we ran into the philosophical puzzlement. I then summarize my account that resolves the problem of mental responsibility. Finally, I turn to an objection, coming from mental compatibilism, that states that I did not really solve the problem at all. The objection will be met with a clarification and an account of mental agency that I provide in the next chapter.

The central mistake that led us to our philosophical puzzlement was to assume that in order to understand every aspect of mental responsibility we need to assume that we are in control of our attitudes. Yet – and this leads me to my diagnosis – It is understandable why we made this mistake. For the problem of mental responsibility *originally* got a grip on us because we assumed that we are responsible only for what we control voluntarily (cf. chapters 0.1, 2.1). Having (indirect) voluntary control over an attitude, however, is merely a necessary condition for *passionate* blaming responses, like indignation, resentment, and guilt, as responses to this attitude. *With this concept of blame in mind*, we turned to the question of how we can be responsible to the standards of rationality. Voluntary control – whether direct or indirect – did not provide us with an explanation (cf. chapters 2.2, 2.4, 4.4).

Because we thought of blame as involving passionate emotions, we suspected that being blameworthy for and responsible to the standards of rationality would require that we are, in some way, in control of our attitudes as well (cf. the second claim of problem presented in

²⁷ I bracket 'their' for the reasons given in subchapter 10.3.

chapter 2.5). However, it was a mistake to think of blame merely as involving passionate emotions. Without justification, we took this narrow concept of blame (blame_H) out of the conceptual space where it is at home – the domain of the voluntariness and control – and tried to force it to live in a hostile environment – in the domain of rationality. Yet being subject to the will and under our control is one thing, while being part of the space of reasons is another.²⁸ Our attitudes are within the space of reasons without always being under our control. Consequently, the way we blame and criticize one another for non-compliance with the standards of rationality must also be different from the way we blame and criticize one another for violating requirements that presuppose voluntary control. There are two faces of responsibility for attitudes.

My solution is thus a *hybrid account of mental responsibility*: On the one hand, we are indirectly responsible for our attitudes, and this form of responsibility can make us legitimate subjects of passionate blame for our attitudes. *The easy solution* provides a satisfying account of the nature of this indirect responsibility. On the other hand, we are directly answerable for our attitudes, and this too is a way of being responsible for them. *The easy solution* does not provide an explanation of our direct answerability. However, answerability is merely being liable to inquiries for one's object-given reasons for attitudes. If our attitudes are well-supported by reasons, they can be evaluated as rational, if they are poorly supported by reasons, they can be considered as irrational. These kinds of evaluation, as all as their normative consequences for our interpersonal relationships, do not presuppose control at all. I thus combine *The easy solution* (under a certain reading of 'responsible') and *mental involuntarism*, thereby making mental responsibility intelligible.

Thus, here is my account:

We are responsible for an attitude if and only if

- (a) we are historically responsible for it in virtue of indirect control, or
- (b) we are directly responsible for it in virtue of being directly answerable for it, i.e., in virtue of our liability to inquiries for our object-given reasons for the attitude.

Obviously, the 'or' is not meant to be logically exclusive. Rather, we can be both indirectly responsible *and* answerable for an attitude. Acknowledging the distinctness of the two kinds of responsibility (as defended in chapter 10), and the kinds of blaming responses they can give

²⁸ This is the central point of Owens (2000).

rise to (as spelled out in chapter 11), is the key to seeing how this very simple account of responsibility allows us to resolve the puzzle of mental responsibility.²⁹

Let me finally point out why alternative reactions to the problem of mental responsibility (as spelled out in chapter 2.6) fail. I have argued against indirect mental voluntarism on the grounds that it provides an extensionally inadequate account of mental responsibility (2.2, 4.4): it ignores our direct responsibility to the standards of rationality (as defended also in chapters 10 and 11). Indirect mental voluntarism thus threatens to collapse into mental nihilism or into anti-responsibilism. While I regard nihilism as a last resort, my replies to the arguments from part II (in chapters 4.3 and 6.4), as well as my argument for direct mental responsibility in chapters 10 and 11, provide us with decisive reasons for rejecting not only nihilism, but also anti-responsibilism. As already pointed out in 4.1.3, anti-responsibilists have to deny the very minimal connection between reasons and blameworthiness I defended in chapters 4.1 and 4.2. Furthermore, I have rejected the direct voluntarist's explanation of mental responsibility by arguing at length against the idea that we can believe for practical reasons (part III).

Here is the objection from mental compatibilism, which I did not consider in detail. Mental compatibilists claim that responsibility for attitudes can only be explained by reference to a form of direct, non-voluntary control that we exercise in believing, desiring, feeling, or intending. I have not argued that we do not exercise such control. I even granted that attitudes are responsive to reasons. Why does this not imply that we exercise non-voluntary control in believing, desiring, feeling, or intending? Thus, so the mental compatibilist complains, I cannot just endorse involuntarism. For in order to establish involuntarism, I have to justify the claim that attitudes are not themselves exercises of control.

The first part of my reply is that involuntarism does not claim that attitudes are not exercises of control *in some sense*. Rather, it claims *that we need not assume that they are exercises of control in order to make it intelligible how we can be responsible for our attitudes*. Involuntarism is a way of *understanding* how we can be responsible for our attitudes. In order to make mental responsibility *intelligible*, we merely need to understand the nature of answerability for our attitudes in terms of the blaming-responses that are made appropriate by non-compliance with rational standards. As soon as we understand the nature of these responses,

²⁹ Though this account is very simple, it gives rise to many further questions the answer to which might not be as simple. First, spelling out the conditions under which an attitude is subject to indirect control might turn out to be quite complex (for an account that puts more effort into this task than I did, cf. Peels 2017). Secondly, giving an account of all the forms of answerability-blame in more detail than I did – e.g., getting clear about the nature of epistemic blame – seems to be a mammoth task insofar as our reactions towards people's attitudes seem to be highly diverse. I return to this latter issue in the Conclusion at the end of the book.

we see that their appropriateness merely presupposes that a subject's attitudes are responsive to reasons in such a way that the subject is answerable for them. By explicating the concept of answerability in terms of these attitudinal responses, I made it plausible that these responses do not presuppose control in any substantial sense (i.e., a sense of 'direct control' that means more than that we are answerable): our relationships to other persons (or to ourselves) can be impaired just because they (or we) *have* certain attitudes. Whether answerability implies that we are, in some *robust* sense, in control of what we are answerable for is a question that need not be decided in order to understand how we are responsible for our attitudes. It is a further question we might have to ask in other contexts, but we need not do so in the context of understanding mental responsibility.

The second part of my reply is that responsiveness to reasons implies a form of indirect control over attitudes. It is true that we have a *special kind of agentive relation towards our own attitudes* – we can exercise our agency with respect to our own mind in a way we could never exercise our agency with respect to the mind of another person. By providing an account of this agentive relation in the last chapter without committing to non-voluntary control, I take away an important motivation for the idea that responsiveness to reasons implies non-voluntary control. Philosophers who try to explicate the kind of control that is implied by an attitude's reasons-responsiveness tend to confuse this special form of indirect control with a form of direct control.

However, I do not wish to deny that there might be other puzzles besides the problem of mental responsibility – puzzles stemming from the very nature of agency – that might require us to make sense of the idea that we are, in some sense to be specified, in direct, non-voluntary control of our attitudes (cf. Hieronymi ms). Yet the assumption of such a form of control is not necessary for making mental responsibility intelligible.

Chapter 12

Responsibility for Ourselves: Sketching an Account of Self-Control

We now lack an important motivation for assuming that we have direct control over our attitudes in any *substantial* sense, i.e., any sense of ‘direct control over attitudes’ that means something *in addition* to saying that attitudes are responsive to object-given reasons. For there is nothing for direct control left to explain: no form of mental responsibility that seems to conceptually demand direct control. *Historical responsibility* requires merely indirect control. And understanding how we are *answerable* for our attitudes does not require the assumption that we are in direct control of our attitudes. It just requires that they are responsive to reasons, and thus capable of reflecting something about ourselves in a way our headaches do not.

However, the conception of mental responsibility spelled out in this book raises another worry: we might feel that our active relation towards our own attitudes is somehow *special* compared to our active relation towards other things in our environment, and that this cannot be captured by assuming that we have merely indirect control over our attitudes.

The intuition about a special agentic relation we have towards our own mind was already put forward in the Stoic tradition. For example, Epictetus claims that anything external to our own mind is ‘none of our business,’ and that the only thing we *really* control are our own judgments about things, rather than the things themselves (*Encheiridion*, art. 1). This claim also excludes our bodily movements from things that are under our direct control. I think the intuition itself does not justify such a strong claim – the Stoics just interpreted the import of the intuition in a very *specific* way. Yet it seems that this intuition about a special agentic relation must be accounted for by considering special forms of control we can exercise only over our own attitudes, but not over mere physical features or objects that are not responsive to reasons, or over the attitudes of others.

Contemporary philosophers think that we can account for this intuition only by spelling out notions of direct control over attitudes. For example, Hieronymi argues that we need to assume a form of control (‘evaluative control’) ‘that prevents its exercise from being grouped with those things that merely happen to one and prevents its outputs [...] from being grouped with those things that one can affect only by acting upon them’ (2009, 147). Similarly, Smith writes that any account that only allows for indirect control over our attitudes ‘obscures the

special nature of our relation to our own attitudes: We are not merely producers of our attitudes, or even guardians over them' (Smith 2005, 251).¹ Moran (2001) thinks that there is an important asymmetry in the way we come to know our own mind and the mind of others, and regards the first form as a special way of exercising agency over one's mind.

I have argued throughout chapters 10 and 11 that we are directly responsible for our attitudes in virtue of being directly answerable for them. Will this not settle the worry these philosophers seem have with mere indirect control over our mind? I do not think so. For part of the worry these philosophers express seems legitimate even if we grant that there is direct responsibility for attitudes. After all, it seems that we relate differently as agents to our own attitudes than we relate as agents to our own headaches, or to ordinary objects in our environment, or towards the attitudes of other people. Since the form of direct responsibility for attitudes I have defended does not imply direct control, we might still be worried that our mind is just one object among others when it comes to how we are in control of it.

In order to meet this worry, I will argue in this chapter that we can do justice to our intuitions about a special form of first-personal agency which we can only exercise over our attitudes even if we grant that the only control we have over our mind is indirect. This is possible because this kind of agency does not take the form these philosophers thought it takes – that is, it does not take the form of direct control over attitudes.

12.1 THE OBJECTION FROM MENTAL AGENCY

If the only control we have over our mental life is indirect – i.e., if we can merely influence our attitudes through our actions –, then our active relation towards *our* mind does not seem to be any different from our active relation towards the mind of *others*, or different from our active relation to any other object in our environment over which we can exercise our manipulative powers. However, this seems counterintuitive. Our beliefs and intentions are not just brute occurrences like some physical pain we suffer, and which we can control by indirect means (e.g., by taking pain killers). Rather, they are directly *responsive to reasons*. As objects in the space of reasons, we expect ourselves to be more actively related to them than to brute physical pain

¹ Smith continues: 'We are, first and foremost, inhabitants of them.' I take the central worry here to be that we are in a special way directly accountable for our attitudes because they are at the center of what we are (our character, or 'quality of will'). Yet Smith's remarks about the supposed insufficiency of a picture of mental control and responsibility that assumes merely indirect control can be read as nicely phrasing the worry that it obscures our relation towards our attitudes as agents. This is the worry I discuss here.

or other phenomena that are unresponsive to reasons, but which are sometimes controllable by indirect means.

The challenge I set out to meet in this chapter is to show how the following claims can be reconciled with each other:

- (1) We can control our mind, and are responsible for it, in a way we cannot in principle control or be responsible for the mind of others or for states that are not responsive to reasons.
- (2) The only control we have over and responsibility we have for our own mind is indirect.

I will argue that we can endorse both claims without committing ourselves to an incoherence. Rather, we can do justice to our intuition about first-personal agency without committing to the existence of non-voluntary forms of control. The key to this reconciliation lies in the acknowledgment of *deliberating* and *reasoning* as genuine actions we can perform for practical reasons, and which grant us a special form of indirect control over our mind – a control we could never exercise over the mind of others. I will spell out this notion of control by building critically on John Broome’s account of reasoning as an activity.²

It became quite common again in recent epistemology and moral psychology to try to account for the distinctive perspective of the agent. Next to theories of responsibility and agency, there are also *theories of self-knowledge* that model our epistemic access to our *own* attitudes as somehow special compared to the epistemic access we have to the mind of others (e.g., by observing their behavior). By contrast, theories that see our self-knowledge as analogous to our knowledge of the attitudes of others are viewed as deficient. Thus, Richard Moran criticizes these theories for ‘either describing something that could just as well be a third-person phenomenon, or transposing an essentially third-person situation to some kind of mental interior’ (2001, 2).³ This is an interesting shift of thought we can observe throughout the last decades in philosophy, and which might have gained some motivation from debates in the philosophy of mind. For after the decline of Cartesian models of the mind, which is evident by the attempts to naturalize the mind and to demystify the perspective of the first person, we can now

² The special responsibility we have in virtue of the specific form of indirect control over our attitudes that I present in this chapter is a special form of historical responsibility. Thus, even if we were to deny that answerability is a form of responsibility for our attitudes, we could still make sense of the intuition that we are responsible for our own attitudes in a way we are not responsible for the attitudes of others.

³ Moran here refers to the idea of a Cartesian ‘internal theater,’ where persons have a quasi-perceptual and infallible access to their own mental life (their ‘experiences,’ desires, beliefs, and so on), which is construed in analogy to our epistemic access to our environment via ordinary perception (only that it is infallible because ‘direct’). For influential discussions against such a model of self-knowledge, see esp. Shoemaker (1963).

witness a revival of accounts that strive for an understanding of phenomena that seem to be absent from an all-too-brute naturalistic view of the world, or a ‘view from nowhere’ (Nagel 1986). We might view this shift of thought as a tendency towards a careful ‘re-enchantment of nature’ (McDowell 1994, Lecture IV), i.e., a tendency that once again seeks to genuinely accommodate phenomena into our ontology that seemed to be worthy of either elimination or reductive explanation.⁴ The phenomenon of a distinctive relation towards ourselves – either epistemically or agentic – might be viewed as one of these naturalistically problematic phenomena.

However, I will not concern myself in this chapter of how this distinctive relation towards ourselves fits within our understanding of our world as a natural world. Rather, I will just assume that it somehow fits into an appropriately naturalistic picture of our world’s ontology – that is, at least, an ontology that includes genuinely first-personal phenomena (see footnote 4 above). My task here is to spell out this *distinctive agentic relation* we have towards our own mind without committing myself to anything that does not fit well within my theory of normativity and responsibility spelled out in the previous chapters. Yet what I do in what follows can be viewed as a contribution to re-enchanting the natural world in a naturalistically unproblematic way: providing space for the idea that we are in some way specially related to our own attitudes without assuming anything like a Cartesian theatre, or an epistemically privileged access to our own mind.

12.2 GETTING A GRIP ON THE OBJECTION

It is not quite clear yet what a satisfying answer to the present objection might be, because, I take it, the objection is not yet formulated clearly. The present subchapter provides a clearer grip on the objection and on what kind of answer we need in order to meet it.

It is helpful to first get a grip about the *unclarity* of the objection itself as it stands. The objection claims that we are responsible for our attitudes in a way we are not responsible for the attitudes of others (or, for that matter, for other brute occurrences or facts that are unresponsive to reasons). It argues that an approach that denies direct mental responsibility cannot account for this *special* kind of responsibility we have for our own attitudes – or, as I will say in what follows, the responsibility we have for *ourselves*.

⁴ Though McDowell uses this phrase as something he wants to distance himself from, and he thus uses it slightly differently from what I say here, I think it is a nice metaphor for philosophy’s genuine rediscovery of ‘problematic’ phenomena like intentionality, qualia, value, or reasons as being a genuine part of our natural world.

However, not any kind of ‘special’ responsibility we have for features of ourselves will justify the claim that this responsibility cannot be accounted for by reference to indirect control. For there might be special responsibilities we have for features of ourselves which *can* be explained by reference to indirect control. Take, for example, the case of responsibility *for our own health*. It seems that we have a *special* kind of responsibility for our own health because we are in a special position to take care of it by indirect means, like doing our daily exercises. There is no one else who can exercise daily for *me* in order to keep *me* in shape. This, however, is a mere contingent matter. We could conceive of scenarios in which someone else’s exercising has – via some futuristic mechanism – positive influence on my own health. In this case, someone else can exercise for me, and this person might, in some sense, be in the same active and responsible position towards my health as I am with respect to my health. Nevertheless, both this person as well as myself can only *indirectly* influence my health – sadly, we cannot be healthy ‘at will.’ Thus, the ‘specialness’ of the responsibility I have for my health is not sufficient to justify the claim that indirect control cannot explain this special responsibility. It obviously can.⁵

I thus propose to understand the objection as follows. First, the special responsibility for ourselves is supposed to be special in a *non-contingent* way. There is no futuristic scenario conceivable in which someone else has the same special kind of responsibility for my own mind as I do. This provides a plausible reading of the objection. For it is hard to see how there can be a form of indirect control which I can exercise over my mind that is not, in principle, exercisable by someone else. It seems that, if I can influence my attitudes by performing an action, then another person could – in principle – perform the same kind of action that has the same influence on my mind. Thus, it is not clear how indirect control could ever explain why we are *non-contingently* responsible for our own attitudes in a way no one else is – simply because these attitudes are *ours*.

However, merely identifying a non-contingent form of control which we can only exercise over our own mind will not do.⁶ For we can easily stipulate forms of control that are, per definition, only exercisable over a certain feature of ourselves. For example, we could understand ‘keeping myself healthy’ as conceptually implying that *I* am doing something that keeps myself healthy. We can make this use explicit by saying that ‘keeping myself healthy’ in this sense means ‘keeping myself healthy by exercising my own agency.’ This would mean that no

⁵ I am grateful to Gerhard Ernst for pointing this out to me.

⁶ I am grateful to Dorothee Bleisch and Konstantin Weber for pointing this out to me.

one else could possibly perform this action for me – not even if this person was wired-up to some futuristic mechanism and causes my health as a result of exercising *their* own agency. Thus, it would be true in this stipulated sense of the phrase that I have, as a conceptual necessity, a control over my own health that no one else has: only I can [keep myself healthy by exercising my own agency.]

Thus, the challenge cannot merely consist in identifying a form of indirect control we can conceptually only exercise over our own mind. For we can do this very trivially as exemplified above. We could just define an action as [causing an attitude by exercising my own agency]. For conceptual reasons, no one else besides myself could possibly perform this action. But the fact that I can perform this action does not mark my own attitudes as anything over which I have a special form of control – a form of control I could not possibly have over mere physical features (i.e., things that are unresponsive to reasons), or over the attitudes of others.

However, we can easily accommodate this insight into our challenge. What we must find is a form of control we can exercise only over our own mind for conceptual reasons, and the reason why we can exercise this control only over our own mind has to be *non-trivial*. Clearly, the explanation of why no one else can [keep myself healthy by exercising my own agency] is trivial. The explanation is the trivial fact that no one else can exercise *my* agency (but only their own). I thus suggest that the challenge is to identify an action that we can, for conceptual reasons, only exercise over our own mind, *while the explanation of why this is so is not merely that we cannot exercise someone else's agency*. I will explain below in this subchapter why I think that my answer to the present challenge meets the criterion of providing a non-trivial explanation.

We now have a clearer grip on the challenge. One way to react to it is to argue that we are not responsible for our own attitudes in any non-contingently special way (that is not trivial). This is not the strategy I will pursue. Rather, I will identify a form of indirect control that is indeed (and non-trivially) only exercisable over our own mind *for conceptual reasons*. Thus, I meet the challenge of accounting for our intuition about a special way of controlling only our own mind by arguing that we have indeed a special way to take charge of only our own mind. We merely need to be careful not to think that this kind of control is fundamentally different from indirect control – rather, it is a *form* of indirect control. Smith is right when she argues that we are indeed, in some sense, more than mere producers or guardians over our attitudes (see Smith 2005, 251): our control over our attitudes, although it is necessarily indirect, amounts to more than merely *causing* or *managing* our attitudes. We can sometimes determine

our attitudes *in a rational way* by means of the activities of deliberation or reasoning. No one else could determine our attitudes in this way.

What do I mean by ‘forming our attitudes in a rational way’? Spelling this out in such a way that we can reply to our challenge will be the main task of the present chapter. However, to get a first grip on the notion, consider the following scenario. Imagine you are discussing with someone about any arbitrary topic. Soon you realize that you disagree with each other. The person you are discussing with now cites reasons in order to convince you of her view. In doing this, she does not aim at merely *causing* a belief in you – otherwise she could as well stimulate your brain in such a way that you thereby come to have the belief desired by her. What she rather wants is that you acknowledge the reasons as reasons and that by comprehending her reasoning, you thereby revise your beliefs *on your own*. This is, for conceptual reasons, no process *she* could possibly cause in you directly (i.e., without *you* being active in changing your view on the matter).

Nobody else could possibly perform the task of rationally determining your attitudes for you by reasoning to them. As Philippa Foot notes at one point, ‘evidence is not a sort of medicine which is taken in the hope that it will work. [...] When given good evidence, it is one’s business to act on it, not to hang around waiting for the right state of mind’ (1958, 509). You are expected to *do* something with the evidence that was given to you in the discussion: to consider it, acknowledge it as reason to believe, and to thereby determine your belief *in a rational way*.

When I say, ‘in a rational way,’ I mean to contrast it with *arational* ways, for example, with causing a belief in yourself, e.g., by taking a pill or by stimulating your own brain with futuristic devices. Note what a self-manipulation of this kind would mean to the person who provided you with evidence: it would show her that, rather than considering the evidence and taking her testimony seriously as something that gives you a reason to change your own view, you instead just *cause* the belief in yourself independently of whether the content of what she said gave you a reason. You might manipulate yourself to believe her from the motive that you want the discussion to come to an end, and because you do not really care about what is true. Surely, this is not what the other discussant intended to happen by giving you reasons to believe. What she aimed at was that you change your belief *in a rational way*. It is important to see that this is different from *merely causing* of your own belief.

I think it is easy to see that the explanation of why we cannot determine another’s attitudes in a rational way meets my criterion of *non-triviality* explained above. To rationally determine one’s attitudes, it is necessary that one reasons to them, which implies that one

exercises one's own agency. However, exercising one's own agency in deliberation is *not sufficient* for determining one's attitudes in a rational way. Rather, the deliberation must also *cause our attitudes non-deviantly*, as will become clear in what follows. There is no possible distinction between deviant and non-deviant causation in the case of causing non-reasons-responsive features like our own health. The explanation of why no one else can determine our attitudes in a rational way is thus not merely the trivial explanation that no one else can exercise our agency, but rather that no one else can cause our attitudes non-deviantly. For other agents, *our* attitudes are just like other physical features in this world. They can influence and manipulate them, but they cannot *reason* to them. Identifying this form of indirect mental agency – non-deviant causation through reasoning – thus meets the challenge of identifying a form of indirect agency that we can non-trivially exercise only over our own mind.

12.3 THE BASIC IDEA: REASONING, CAUSING, AND RATIONAL DETERMINATION

Reasoning is something we do. It is a mental activity of ours that can bring us to satisfy some of the requirements of rationality (Broome 2013, 3).

One might read Broome here as saying something controversial. However, as I will try to make clear in this section, there is something wholly uncontroversial about his claim that reasoning is an activity. Nevertheless, we should refrain from endorsing some of the details of his account of reasoning as an activity, to which I will return below. Yet the general idea that reasoning is a special kind of activity by means of which we can control our attitudes will help us to identify the special agential relation towards ourselves.

After I clarified the sense in which reasoning is an activity, especially by distinguishing it from another, passive phenomenon which I will call 'coming to a conclusion,' I will argue that there is a way in which we can determine our own mental life by means of reasoning in a way no one else could: we can sometimes non-deviantly cause our attitudes by reasoning to them, i.e., we can determine them *in a rational way*. I will leave a discussion of Broome's claim that we can satisfy requirements of rationality through reasoning to section 12.5. The present focus is on reasoning as an action.

As I have argued in chapter 8.2, there are two senses in which we can 'form a belief.' When we read a newspaper and, as a result, come to believe that a certain event *e* happened, then we (a) performed the activity of reading a newspaper and (b) our doxastic state changes from [not-believing that *e* happened] to [believing that *e* happened]. Thus, when we say that someone formed a belief *by* reading a newspaper, then this phrase is ambiguous. For the *by*-

relation might either express a relation of *identity* between the actions of *forming a belief* and *reading a newspaper*; or it might express a relation between *an event of doxastic change that was caused by reading a newspaper* and the action of *reading a newspaper*. I have argued in chapter 8.2 that this event of doxastic change that results from our reading the newspaper is not active in any substantial sense. We are thus only in direct control of what we do in order to bring about doxastic changes. We are not, however, in direct control of these doxastic changes themselves.

We can apply this general distinction to reasoning. Imagine someone wants to solve a mathematical problem. This person might then sit down, write mathematical symbols, think about what symbols to write next, and finally come to a solution of the problem. We might say that what the person *does* is reasoning about the solution to the problem. This reasoning of hers is made up of the activities of *writing* and *thinking*.⁷ Since writing and thinking are activities, reasoning in this sense is an activity. Thus, Broome's claim that reasoning is an activity seems to be trivially true.

However, it is not quite straightforward whether Broome uses the term 'reasoning' in this way. Rather, he also seems to have in mind to include the passive change of doxastic state – the coming to a conclusion – into the activity of reasoning, and to view this change itself as something that is active:

There are two aspects to reasoning. One is working out a conclusion on the basis of the premises. The other is taking up an attitude towards the conclusion. It is tempting to try and divide the reasoning into two stages according to these two aspects: first the working out of a conclusion, then the acquisition of an attitude (Broome 2013, 243).

According to Broome, we should resist this temptation. In some sense, he seems to claim, we should not distinguish between the activity of reasoning the mathematician is performing and her coming to a conclusion ('taking up an attitude towards the conclusion'). Rather, reasoning and coming to a conclusion ('taking up an attitude,' or 'forming a belief' in the sense of a doxastic change) are one and the same: 'Since the reasoning is an act, and is the forming of a belief, the forming of the belief this way is an act' (235).

However, this is puzzling. *First*, there is a clear sense in which I can reason but not come to a conclusion. In order for our actions to reach their desired end, the world must

⁷ If one doubts that her thinking is an activity, imagine her sitting in her chair, her fist under her chin (like Rodin's *Le Penseur*), maybe sometimes scratching her hair. While her thinking is maybe not strictly identical to these activities, I would argue that these activities are visible manifestations of her activity of thinking (rather than mere consequences of what is going on passively in her head; and rather than mere 'signs' of her thinking).

cooperate. Not coming to this end does not disqualify the activity as the activity it is – unless the activity is one we pick out by using a success term, like ‘scoring for one’s team’ (which is not the case with ‘reasoning’). When we were hiking to Santiago de Compostela and do not arrive, that does not mean that we were not hiking to Santiago de Compostela (although we *did not* hike to Santiago de Compostela). In one important sense, we *were* hiking to Santiago de Compostela, but sadly we did not make it there, and thus did not reach our aim. Similarly, when we are crossing the street and we get hit by a car, that does not mean that we were not crossing the street. In one important sense, we were, for otherwise it would not be intelligible how we could get hit by a car on the street (‘How did you get into such a terrible accident if you were not even crossing the street? Was the driver drunk and hit you on the sidewalk?’).

Thus, ‘hiking to X’ and ‘crossing X’ are sometimes used in such a way that they *have* an aim, but they remain the activities they are even if they do *not* reach their aim. These terms are thus no success terms which imply the reaching of their aim. By contrast, if we score a point for our team, that means that our action was successful: we cannot score a point for our team but, alas, we have failed to hit the basket (‘scoring a point for our team’ is thus a success term). ‘Reasoning’ does not belong to the latter category of words, but to the former. Our mathematician would still count as having reasoned if she would not have found a solution to the problem that concerned her. Maybe she was too tired to find the solution, and thus her own condition hindered her reasoning to come to its intended aim. But she was reasoning, and it can even be said that she *was reasoning to a solution* of her problem (however, she *did not* reason to a solution).

Secondly, it is not at all obvious why we *should not* draw the distinction between the activity of reasoning and the coming to a conclusion. In principle, we are allowed to draw any distinction in language that seems to be useful for our purposes. Noting the difference between the activity of reasoning and the coming to a conclusion that results from it is helpful in order to see what we can control directly (the activity) and what we cannot control directly (the coming to a conclusion). When Broome argues that ‘[s]ince the reasoning is an act, and is the forming of a belief, the forming of the belief this way is an act’ (235), he has in mind that reasoning is *the coming to a conclusion* – this latter event is what he means by ‘forming of the belief.’ Otherwise his argument would be trivial (it would just state that reasoning is an act, and since this act is the act of reasoning, reasoning is an act). But reasoning, understood as an action, is precisely *not* the coming to a conclusion, but rather the activity that *precedes* it. Broome here falls prey to the ambiguity in ‘forming a belief’ I spelled out in chapter 8.2, and which I re-stated above. Since reasoning is precisely *not* the forming of a belief in the relevant

sense (coming to a conclusion), Broome's simple argument for the claim that 'forming a belief' (in the sense of 'coming to a conclusion') is an action is not sound.

Broome could reply that he *wants* to use the term 'reasoning' as a success term, so that every piece of active reasoning is only reasoning, in this new stipulated sense, insofar as it involves the coming to a conclusion. However, this does not warrant his point that coming to a conclusion is itself an action. It is just introducing a term of art that encompasses an action as well as its result – like the term 'scoring for one's team.' Introducing this new sense of 'reasoning' does not magically transform the result of reasoning into an action – as introducing the term 'treefloor' for any objects that are constituted by a tree and a floor does not transform the floor that is part of this complex object into a tree.⁸ Just because we can actively 'reason' (in Broome's success term sense) does not mean that our coming to a conclusion, which is part of reasoning in this sense, is itself something we did. Compare the action of 'scoring for one's team,' which has as its result that the ball is falling through the opponent's basket. The ball's falling through the basket is part of scoring for one's team. But the ball's falling through the basket is not something a person *did*.

However, I still think there is a true core in Broome's idea that reasoning constitutively involves coming to a conclusion. It is not that in order to count as having reasoned, we must come to a conclusion. Rather, in order to count as having *determined our attitudes in a rational way*, we must come to a conclusion (because it is a success term).⁹ This is a special *type* of reasoning. To better get this type into view, consider first what Broome argues is amiss in a what he takes to be a too simplistic picture of reasoning that ignores the close connection between the activity of reasoning and the coming to a conclusion. He describes this picture as follows:

[I]n reasoning you call to mind some of the premises, and doing so jogs into operation an automatic process that causes you to acquire a conclusion-attitude. [...] On the jogging account, reasoning is a bit like listening. Hearing is an automatic process through which you can acquire some beliefs[...].

⁸ I owe the 'treefloor' example to Gerhard Ernst, who brought it up in other philosophical contexts.

⁹ One might object that this will depend on whether we use the term in past progressive or simple past. For to say that 'I was determining my attitudes in a rational way when you interrupted me before I finished' is grammatically unobjectionable, while 'I determined my attitudes in a rational way when you interrupted me before I finished' does not sound correct. I have no objection against the first use, of course. It is just important to keep in mind that using the past progressive is just to say that I was engaged in the activity of reasoning. When I say that someone determined their attitude in a rational way (by reasoning to it) I mean that this person was engaged in a complex activity of reasoning that resulted in its constitutive aim of coming to a conclusion. I will just avoid using the past progressive in order not to confuse.

Listening is something you can do to jog your hearing into operation on a particular occasion. (Broome 2013, 226)

According to the jogging account of reasoning, the activity of reasoning consists of (a) performing certain actions characteristic of reasoning, which then (b) cause a belief. Everything we do when we reason would be activities like, for example, concentrating on certain facts. Whether we form an attitude as a result of these activities is, to use a phrase from Galen Strawson, a matter of our ‘mental ballistics’ (Strawson 2003). After we concentrated on our premises, there is nothing more to *do* than to hope that this will cause the correct belief in us (see chapter 2.1). Am I sketching an all-too-simplistic account of reasoning along the lines of the jogging-account here?

However, the jogging-account is at best an unfavorable caricature of what I wish to say. First, the actions that constitute reasoning are much more complex and multifarious than mere attendance to certain facts. Broome himself later adds activities like ‘deriv[ing] some beliefs, hold[ing] their contents in your mind, bring[ing] the contents of other beliefs to bear on them, and so on’ (2013, 231). What is remarkable about these activities is that they do not provide us with *new* evidence. Rather, in performing these activities, we are dealing with the evidence *we already possess*. By attending to certain facts and trying to remember something, we do not learn anything new, but rather we aim at drawing the right conclusions from what we already know.¹⁰ In this way, reasoning differs from listening: we listen in order to hear something we do not yet know, or something we would not be in a position to know without listening carefully. This is, to make a more general point, the difference between reasoning and sense perception: while the latter provides us with new evidence, the former organizes the evidence we already have.

Secondly, reasoning is *sometimes* different from listening in that we do not merely *cause* our attitudes by reasoning to them (as when we do when we listen to something and thereby cause ourselves to have a new belief). Rather, we sometimes determine them *in a rational way* by reasoning to them. Broome does not see how the jogging account can amount to any more than an understanding of reasoning as mere causation of an attitude through previous actions. According to the jogging account, even the following kind of process might count as reasoning:

¹⁰ However, through reasoning we might learn something new about the relation between our attitudes, for example, that we have contradictory beliefs. The claim must thus be specified in the following way: through reasoning, we do not learn anything new besides facts about our own mental constellation. Thanks to Dorothee Bleisch for pointing this out to me.

[S]uppose [reasoning] sometimes goes wrong because of a quirk of your psychology. For instance, when you call to mind that it is raining and that if it is raining the snow will melt, this jogs into operation an automatic process that causes you to believe you hear trumpets. (Broome 2013, 226)

Broome views this as a *reductio ad absurdum* of the jogging account. According to this argument, reasoning never *merely causes* a belief. In order for something to count as reasoning, it has to cause the belief *in the right way* (what I call ‘a rational way’). We can even conceive of cases of deviant causation which present the jogging account with a challenge similar to Donald Davidson’s challenge of finding the right kind of causal process that explains the relation between a reason for an action and the action that was caused by that reason (or thoughts about that reason). In such a scenario, my calling to mind that it is raining and that if it is raining the snow will melt would cause me to believe that the snow will melt. However, this process would still not count as reasoning, according to Broome’s understanding, because my belief might have been caused by some odd causal chain. For example, my calling to mind the premises might have caused me to have a serious defect in my brain which then accidentally resulted in my believing that the snow will melt. For sure, Broome would say, such a process cannot count as reasoning.

However, as I noted above, it is possible that you engage in reasoning but that you get disturbed before you are done. The disturbance might be someone screaming outside on the streets, or it might be a malfunction in your own brain that causes you to end up with an odd belief, or even with the belief that is logically implied by the premises you considered. It does not matter what it is that disturbs you, or what the results of this disturbance are: your activity that was interrupted does not cease to be an instance of reasoning just because it was interrupted, or just because it deviantly caused some mental state. You just did not reach a conclusion (in a rational way). Even if the causal chain that lead to your conclusion-attitude was deviant, you might have reasoned beforehand when you were focusing on the premise-attitudes.¹¹

Nevertheless, I wish to defend a similar claim as Broome does: there is an important difference between an event of a belief being merely caused by a causal process, and the event of determining one’s belief in a rational way. Though both events might be preceded by an activity of reasoning, only the latter event ensures that I am responsible for the belief resulting

¹¹ Note that I am not claiming that focusing on some thoughts is sufficient for reasoning. At least you have to focus on these thoughts with the (implicit or explicit) aim of drawing a conclusion from them: you have to believe that what you focus on is somehow relevant to a question you pose yourself.

from my reasoning in a way nobody else could possibly be responsible for it. In order to meet the challenge presented by Smith, it is enough to see *that* there is a difference between merely causing my belief and causing it in a rational way. The details of this distinction do not matter. Spelling out the distinction would amount to meeting a similar challenge as the challenge from deviant causation in action theory. Instead of meeting this further challenge, I will here focus on spelling out how the fact that there is a distinction between *merely causing* and *rationaly determining* one's attitudes will imply a conceptual asymmetry in the control we have over ourselves and the control we have over others. We will see this conceptual asymmetry by considering thought-experiments. These experiments focus on the attitude of belief. I will consider whether we can rationally determine our emotions (or what it would mean *to reason to an emotion*) in section 12.5.3.

12.4 REPLYING TO THE OBJECTION: A CONCEPTUAL ASYMMETRY

Can we rationally determine someone else's belief? In order to meet the worry about first-personal agency, I must argue that we cannot conceive of a scenario in which a person rationally determines someone else's belief. My argument will proceed by considering what I take to be the most plausible scenarios that could be thought of as scenarios in which someone rationally determines someone else's belief. I will argue that in each of these scenarios, it is either not the *other* person who rationally determines the other's belief, or the other's belief is not rationally determined at all.

If I reason aloud and someone else follows my reasoning, this person might come to have a belief as a result of my reasoning. Does this amount to rationally determining the other person's belief? No, because the person either reasoned to her belief by herself, and thus determined it on her own, or if she didn't, then my reasoning either merely caused her belief or it is not the primary causal explanation of how her belief came about. These are the three possible scenarios I will comment in what follows.

In the first mentioned scenario, my reasoning motivates the other person to follow it and to have her *own* chain of reasoning. Her own reasoning might then result in a rational determination of the belief, and my reasoning was only an *indirect* cause of this determination: for it first caused a chain of reasoning in the other person which then determined her belief.

In the second mentioned scenario, my reasoning *directly* caused her belief without there being an intermediate chain of reasoning of the person I 'discussed' with (there is no real *discussion* when someone does not follow the other person's reasoning). This scenario seems

unrealistic, but it is not conceptually impossible. Imagine that the other person did not listen to me, but that the sound of my reasoning aloud caused her brain by some odd causal accident to be in exactly such a state that she came to believe my conclusion. For sure, this does not count as a rational determination of her belief by herself.

In the third mentioned scenario, my reasoning causes some other event which then merely causes the other person's belief that my conclusion is true. This scenario is structurally analogous to the first one, but it differs in that the second causation involved is a *mere* causation rather than a rational determination by reasoning. One might think of the following futuristic scenario: the person I give reasons to believe that *p* is connected to a machine that causes her to believe that *p*. In this scenario, my reasoning causes the machine to operate, which then causes her to believe my conclusion. This is mere causation of a belief, and no rational determination.

Thus, we cannot imagine the scenario I described as one in which I rationally determine the other person's belief. This is good evidence that it is impossible to determine anyone else's belief in a rational way. However, my point is that it is *conceptually* impossible to determine anyone else's belief in a rational way. Thus, one might think that there are, in principle, other scenarios conceivable in which my reasoning rationally determines someone else's belief. Let us now try to conceive of a science-fiction case in which rational determination of someone else's belief is possible. We will notice that we have a hard time conceiving of these cases in such a way.

Imagine that Tina's and Tom's brains are connected in such a way that whenever Tina reasons to the belief that *p*, Tom also reasons to the belief that *p*. Is Tina able to determine Tom's belief in a rational way by engaging in a reasoning process that takes place not only in her mind, but also in Tom's?

We first should note that the scenario is not sufficiently described in order to answer this question. For it leaves unanswered the question of whether Tom exercises his own agency in reasoning to his belief. In case he does, it might well be that his belief is rationally determined *by him*. But then it seems that he could also *choose* not to reason in this way, and the connection between Tina's and Tom's brain is not as strict as the scenario suggests. Or Tom does *not* exercise his own agency. Then it is as if Tina gained control over Tom's mind. The scenario is similar to one in which a mad scientist is causing a reasoning process in Tom's mind that results in the determination of Tom's belief. Here, not *Tom* determined his own belief – although it might *seem* to him as though as he did. Rather, it was the mad scientist, or Tina,

who both controlled the reasoning process as well as the resulting belief by reasoning to it. Did *the mad scientist*, or *Tina*, bring about the belief in Tom *in a rational way*?

Our intuitions in science-fiction scenarios might be unclear, because we have a hard time imagining the details of the machinery that allows a brain connection or stimulation in such a way. However, if our intuitions are unclear, I might well just point out that my concept of control by rational determination can explain the special kind of control we have over our own mind as long as we cannot imagine the details of such a case. Maybe we can do so in the future. However, this might fundamentally challenge and change our views of the rational self and personal identity. My task here is not to challenge these views by new empirical findings, but rather to analyze our current conceptual scheme. So, in order to make the objection I discuss here as strong as possible, let us assume for a moment that we have a sufficiently clear way of conceiving of thought experiments of this kind and ask whether the thought experiment under discussion can challenge my claim that nobody else can determine our own attitudes in a rational way.

I think the scientist, or Tina, might be conceived of determining Tom's belief in a rational way in the scenario I described. However, as I noted in the last paragraph, if such a scenario is conceivable, this might challenge our views of ourselves. If another person can rationally determine someone else's belief, this might blur the lines between these two persons. Again, we have to distinguish two scenarios in order to see this more clearly.

First, both Tina and Tom might be exercising their own agency when they reason simultaneously to the same conclusion by the same reasoning-process while their brains are connected. Note the scenario must be such that the reasoning of Tom is caused or necessitated by the fact that Tina is reasoning, so that she gains full control over what Tom does. As noted above, I doubt that Tom can be conceived as reasoning on his own in such a case, for he does no longer have the capacity to choose not to reason. However, Frankfurt-style compatibilists might disagree with me on this point.¹² So let us grant for a moment that it is possible that Tom is still active although Tina's reasoning necessitates his reasoning via the futuristic mechanism. In such a scenario, if we can conceive of it, it seems to me that Tom and Tina might *share one and the same reasoning process*. Remember that the reasoning of both is connected via some futuristic mechanism, and that both persons are, given this interpretation of the scenario, involved in the activity of reasoning. In this case, it is hard to see why we should distinguish

¹² Cf. Frankfurt 1969. For a Frankfurt-style compatibilist account of responsibility and control, cf. Fischer/Ravizza 1998.

between the two persons insofar as they are involved in reasoning. At least in one respect, it seems that their minds have merged. Thus, it is not surprising that Tina can rationally determine Tom's belief, because the identities of Tina and Tom partly merge into one another.

However, take the second possible interpretation of the scenario: Tom is completely passive, and merely experiences his reasoning process as being under his control while being connected to Tina, who is the only one who exercises her agency. Is this a rational determination of Tom's belief by Tina (or, for that matter, a mad scientist)? It is not, because Tom is not *reasoning* at all. Rational determination can only take place as a result of the *activity* of reasoning. But Tom is not active, and thus he is not reasoning in the relevant sense. Thus, there is no rational determination of Tom's belief by someone else.

I have described what I take the most plausible candidates of conceptually possible scenarios in which someone might be conceived of determining someone else's belief in a rational way. All these scenarios, however, turned out to be either scenarios in which people determine their belief on their own (as in the first interpretation of the discussion-case, and in the first version of the Tina and Tom-case, in which two people seem to *share* a reasoning process), or scenarios in which there is no rational determination taking place at all (as in the remaining cases). I thus conclude that I have identified an activity which we can, for conceptual reasons, only exercise over our own mind, but which is nevertheless indirect: it is the activity of reasoning which results in the determination of one's attitudes. Some worries have to be addressed, however.

12.5 REMAINING ISSUES

There are three issues that remain undiscussed until here, but which are either pressing to the reader or which I mentioned before but did not yet discuss. The first, pressing issue, is a further objection one might see: the only thing we directly control is our reasoning, not the determination of our own attitudes; but it is only the latter event that can directly influence only our own mind, while reasoning is an activity that can, in principle, also influence the mind of others; so it might seem that, after all, I did not identify an activity that we can only exercise over our own mind.

The second issue is the question of whether we can control our rationality by reasoning – Broome's claim that I quoted at the beginning of chapter 12.3. If my account cannot account for a connection between rationality and reasoning, then this might be viewed as a defect of it.

The first and the second issue are interrelated, for the question of whether we are active in determining our attitudes will bear on whether or in what way we can control our rationality by the special kind of reasoning that results in the determination of our attitudes.

Finally, I should say something about how we control other attitudes than beliefs, especially intentions and emotions. I will address the issue of controlling intentions while I discuss the second issue about controlling our rationality, and I will turn to emotions separately in a final subsection. Do we have a special kind of control over our intentions and emotions that no one else has? I argue that we do, but that the form of control that we exercise over intentions and emotions by reasoning is more indirect than the control we exercise over our beliefs by reasoning to them. This has some interesting implications for accounts about the nature of emotions.

12.5.1 ACTIVELY DETERMINING ONE'S ATTITUDES

Since rationally determining a belief is the event of doxastic change I called 'coming to a conclusion,' determining one's belief is not itself active. This might sound counter-intuitive and call into question the very idea that, according to my account, there is a special kind of control we can only exercise over our own mind. For self-determination is nothing we control directly, given my commitment to the idea that the only things we control are our actions, but not the non-actions that result from them. The worry is thus that I failed to identify a special form of control we only have over our own mind, but not the mind of others. After all, only our reasoning is active, and we can influence ourselves as well as others by means of our reasoning. But it is, according to what I committed myself to here, precisely *not* under our control whether our reasoning results in the rational determination of a belief. After we are done with our reasoning, all we can do is sit and wait, and hope that it will result in the right conclusion.

On the one hand, I am ready to bite the bullet here and just point out that anything can go wrong after we are done with our active reasoning, but that this does nothing undermine the idea that we have a special form of indirect control over our own mind. On the other, I would like to re-frame the all-too-crude picture provided by the jogging account of reasoning that suggests that our coming to a conclusion is not even indirectly controlled, and that it is a mere matter of luck whether it will take place after our reasoning is done.

To make both points, it will help to compare the activity of reasoning and its result of coming to a conclusion with other activities that have a constitutive aim (but lack a constitutive

result). Take, for example, the activity of searching for X.¹³ When we search for something, it is quite clear that it is not yet determined whether we will find it, although finding is the aim of searching. In principle, even if we are very good searchers, anything could come between the activity of searching and the finding: a heavy storm might urge us to stop our search and look for shelter; or we might get exhausted by our search and interrupt it in order to continue on another day, or even give it up forever because it is too difficult and not worth the effort ('finding the needle in the haystack'). *The very same events or reasons could interrupt our reasoning and thus prevent us from coming to a conclusion*: a heavy storm does not make a good atmosphere for silent contemplation, and reasoning will at some point exhaust us so that we need to get rest, or even so much that we judge that it is not worth the effort (how many mathematicians in history stopped looking for a solution to a problem because they judged that they had more important things to do in life?). My account, which separates reasoning from its result – the rational determination of a belief, or the coming to a conclusion – is doing justice to this obvious analogy between reasoning and searching: there is indeed a bit of luck involved in whether these activities reach their aim.

However, this does not yet meet the objection that I failed to identify a non-contingently special kind of control we only have over ourselves. After all, rational determination or coming to a conclusion is not itself an exercise of control, according to what I say. Again, compare the coming to a conclusion with the finding of an object (i.e., compare the aim of reasoning with the aim of searching). Does the fact that the finding that results from my search is not something I *do*¹⁴ mean that it is not under my control? This only follows if we employ an overdemanding concept of control: finding is under our control only if there is nothing that could have gone wrong, i.e., only if we can find something as soon as we want to find it. For sure, given our ordinary concept of control, there is a lot of stuff in this world we control even though there might have been a disturbance.¹⁵ I control my coffee cup and the way I bring it to my mouth even though it might well have happened that my house exploded before I exercised this capacity of control. Thus, I can exercise indirect control over whether I find something – for example, by making a big effort in searching for it. Thus, I can exercise indirect control over

¹³ In fact, reasoning might be conceived of as a special kind of searching: searching for a conclusion. However, I here have in mind the kinds of searching that not a form of reasoning, like looking for one's keys.

¹⁴ When I am done with searching and I was successful, this just *means* that I found what I was looking for – I do not have to perform the action of *finding* after I am done with searching. There is no such action, for it is the result of searching – as there is no such action of 'forming an attitude' after we are done with reasoning.

¹⁵ This point is made, e.g., by Oakley (1992, 124-127), who follows Stocker (1982, 408-411). For some worries with the idea that we ever *control* our attitudes indirectly, cf. Lindner *fc*. I disagree with Lindner that the kind of influence we can exercise indirectly over our mind does not amount to control.

whether I rationally determine my beliefs – again, for example, by making an effort in reasoning. In this respect, coming to a conclusion is not a mere matter of luck, as finding something is not a mere matter of luck. By making efforts, we can increase our chances of success.

As I explained at the outset of this chapter, the special kind of control over ourselves remains an indirect form of control. That does not mean that it is not a form of control we can exercise only over our own mind. Furthermore, the control we exercise over our own mind is as imperfect as any control we exercise in this world: other events might prevent our activities to achieve their constitutive aims or desired results. Nevertheless, we can be said to control these results by means of our actions. Since, as I have argued, no one else than we ourselves can rationally determine our own attitudes by reasoning to them, we have a special kind of indirect control over our own mind that no one else can exercise over it. This result both *explains* an old philosophical intuition that we stand in a special relation to our own mind, and it *demystifies* this kind of control by reducing it to a familiar form of indirect voluntary control as well as by acknowledging its imperfectness. While the Stoics were right in pointing out that we have a special responsibility for our own mind, they were wrong in assuming that anything beyond our own mind is ‘not our business.’¹⁶ If anything over which we have only imperfect control would not be our business, then nothing in this world would be – not even our own mind.

12.5.2 CONTROLLING ONE’S RATIONALITY THROUGH REASONING

What can we do if we notice that we fail to satisfy a standard of rationality? I think the answer depends on what kind of standard it is. Sometimes, reasoning is an effective means to satisfy the standard. At other times, reasoning will not to. For it will not reliably lead us to the formation of an attitude. I think there is a notable asymmetry between standards of theoretical and standards of practical rationality.

Take, first a standard of theoretical rationality, like the standard not to believe in contradictions. If we notice that two of our beliefs contradict each other, reasoning may indeed bring us to drop at least one of the beliefs and thus help us to become theoretically rational. By merely evaluating and attending to the evidence we already possess, we can see that the two contradictory beliefs cannot both be true, and thus suspend judgment about which of two contradictory beliefs is true instead of believing in both. Sometimes our available evidence will even support only one of the contradictory beliefs and we can then, through considering our

¹⁶ Cf. Epictetus, *Encheiridion*, art. 1.

available evidence, rationally determine which one of the two beliefs is true. Again, it is important to note that engaging in reasoning cannot *guarantee* that we satisfy the standards of theoretical rationality. The world could come between our action of reasoning and its constitutive aim or desired result of coming to a conclusion. Yet, as I have argued, this does not undermine that we control our rationality through reasoning.

What about practical rationality? Take the practical irrationality of weakness of the will. Can reasoning help us to overcome *akrasia*? Is it helpful to *reason* when I am akratically lying in bed, believing that I ought to get up, but not intending to do so? The answer suggests itself: I already *know* what I ought to do, so why on earth should I waste my time with even more reasoning, thus giving in to my *akrasia* by engaging in cognitively wasteful reasoning and continuing not to do what I know I ought to do (getting up)? This point is worth emphasizing: to *reason* when one ought to *get up* instead is to engage in akratic action. Reasoning would be practically irrational.¹⁷

It is a common assumption in contemporary philosophy that we can reason not only to beliefs, but also to *intentions*. Though I do not wish to argue that reasoning to intentions is impossible, I here want to suggest that this is not the *most common* and not always the *most effective* way in which we actively form intentions. When there is no temporal gap between the present and what we ought to do, then we can form the intention to do what we ought to do *by just doing it* (otherwise it could not be something we ought to do *right now*). The best way to get out of bed in the morning when one does not intend to do so is not to reason yourself to the intention to do so, but rather to *get out of bed*. Performing this action, which is under your direct voluntary control (as long as you are not suffering a pathology), will imply that you had the intention to do so at some point before you got up.¹⁸

What if there is a temporal gap between the present and what I think I ought to do? Say, I think I ought to work on a manuscript tomorrow, but I realize that I do not yet intend to work

¹⁷ Reasoning might still be the second-best option in some cases, as when you are trying to remember the reasons for getting up in order to make it easier for you to get up. Furthermore, both reasoning and getting up might sometimes be *sufficiently* supported by reasons, so that you are not blameworthy to any degree if you either reason instead of getting up. In this case, however, reasoning is not what you ought to do. Rather, you ought to [either reason or get up]. When you continue lying in bed instead of getting up, however, further reasoning will become akratic at some point in time, because further reasoning will no longer be one of the best options for you. Thanks to Dorothee Bleisch for pointing out these complications.

¹⁸ It might seem odd to some that by performing the action, one magically creates the fact that one intended to get out of bed at some point in time before one got out of bed. However, this mystery vanishes as soon as we realize that attitudes, like intentions, are ascribed in order to explain behavior. They are no actions we perform (cf. ch. 9.3), but rather entities we postulate in order to understand each other. For how we can make it true that we exercised our agency in deciding by performing the action we decided on, cf. Soteriou *fc*.

on the manuscript tomorrow. This is a form of practical irrationality as well.¹⁹ Again, it does not seem to be the only effective method to *reason* to the intention that you ought to work on the manuscript tomorrow. Considering the reasons in favour of working on the manuscript might confirm your belief that you ought to work on the manuscript, and having present these reasons before your mind might well lead to a rational determination of your intention to work on your manuscript via your practical judgment that you ought to work on the manuscript. But it might well not lead to you having this intention: the world, or your brain and body, might not cooperate with your judgment. So, what should you do in this case? Reasoning does not help, given the description of the case. But that does not leave us without any further method. What exactly will help best is an empirical question, but probably often mindful activities like meditation, or getting some fresh air on a walk, might help one's brain and body to cooperate with one's best judgments again.

Thus, reasoning might indeed help us, as Broome points out, to satisfy some of the standards of rationality. It seems to be especially helpful and effective in the realm of theoretical rationality, because theoretical rationality is concerned with beliefs, rather than intentions. While beliefs are directly determined by reasoning, intentions might not cooperate with our judgments about what we ought to do even though we are settled on what we ought to do. So there seems to be, for conceptual reasons, a further 'barrier' between judgment and intention that might hinder our reasoning to be effective regarding the formation of an intention.²⁰ Furthermore, it is important to keep in mind that since reasoning is itself an action, it can be akratic as well. Reasoning can thus constitute your practical irrationality, rather than remedy it, but it can never constitute your theoretical irrationality (the objects of theoretical irrationality are beliefs, rather than intentions).

However, in principle our reasoning might be ineffective in the theoretical realm as well (we might talk of 'epistemic akrasia' in such a case), so that other, more indirect, means

¹⁹ Pace Kiesewetter (2017, 190-192), who thinks that one is irrational only when the time of action draws near. I argued that this is wrong (cf. ch. 5.1.4).

²⁰ This idea might pose a problem for Owen's (2000; 2017, Introduction) theory of mental control. He argues that we have reflective control over our intentions, but not over our beliefs, by having the ability to make our intentions conform to our practical judgment. This seems to imply that we have somehow *more* control over our intentions than over our beliefs. By contrast, my picture suggests that it is sometimes very hard to make our intentions line up with our judgments, and if they do not do so automatically, we have to retreat to indirect means. There is no other, more direct control, we could exercise over our intentions. However, Owens' theory nicely captures the fact that there is not (normally) a barrier between one's reason for belief and belief: according to him, our evidence usually causes our beliefs directly, and there is nothing but indirect control over our beliefs. I think this might support the idea that reasoning is more often effective in the realm of theoretical rationality than in the realm of practical rationality.

are needed in order to make us theoretically rational. We might think of such indirect means, like meditating or going for a walk, as actions we perform in order to re-charge our disposition towards being rational. These more indirect means seem to be especially required in order to make us practically rational, i.e., to form the right intentions given our beliefs and our evidence about what the world is like, and our evidence about what is right. Though reasoning is a way to satisfy standards of rationality, it is not always an effective one, and there are other, and often better, ways to do so.

12.5.3 CONTROLLING ONE'S EMOTIONS THROUGH REASONING

Can we ensure that we have rational emotions by reasoning to them? The way we control our emotions through reasoning will depend on which theory about the nature of emotions is right. Since I remained neutral on this issue throughout this book, there are at least two different kinds theories to consider. First, the cognitive theory of emotions, which roughly claims that emotions are a kind of belief or judgment (Solomon 2007). Secondly, 'feeling'-theories of emotion à la William James (1884), and perceptual theories of emotion (Price 2015; Tappolet 2000; 2016). I lump together the latter two theories for my purposes here because they agree in that emotions are no cognitive states. If emotions were cognitive states, then everything I said about beliefs can be expected to be roughly applicable to emotions as well. In this case, we could reason to emotions as we could reason to beliefs or judgments.

However, the way our reasoning in fact influences our emotions might question whether cognitive theories of emotions are right. Consider a case of recalcitrant emotion, which are traditionally conceived as irrational emotions: John is afraid of flying, he nevertheless enters a plane because he judges that the flight is not as dangerous as his fear suggests, but he remains afraid throughout the whole flight. Could John *reason* his way out of his emotion? For sure, this is not always possible. And the emotion need not even be pathological: great heights make most people feel dizzy and afraid even though they know that nothing will happen to them (think of someone who jumps out of a plane with his parachute for the fifth time or so). This emotion cannot be dealt with by *reasoning* it away (and if it is not too strong, there is no need to get rid of it at all). However, if emotions were beliefs, then one would expect a more reliable responsiveness to reasoning.

In this way, emotions seem to be more like intentions: our control over them is somehow mediated through our judgments. While we might correctly judge that we ought to do something, or that there is no danger involved in a given situation, our intention to do so might

not follow our judgment, or, respectively, our fear may still persist even in face of our judgment that there is no danger. However, while we can cause our intention by just doing what we think we ought to do, we usually cannot change our fear of something which we judge is not dangerous *by performing an action*. Rather, we often have to live with our emotions throughout our actions, and thus have to ‘overcome our fear,’ which does not mean that we do not experience our fear when we act. This just highlights a difference between intention and emotion: while doing what one thinks one ought to do implies that one intended to do it, overcoming one’s fear by doing what one thinks is not dangerous does not imply that one now has a rational emotion towards one’s action – one might still be afraid despite believing that there is no danger.

Thus, non-cognitive theories of emotion might seem more plausible given the way our emotions respond to our reasoning. If emotions were similar in relevant ways to bodily feelings or to perceptual experiences, then it would be intelligible why they do not respond to reasoning in the way beliefs do. For our bodily feelings of pain do not vanish if we engage in reasoning, nor do our perceptual seemings change when we reason to the belief that things are currently not as they seem (as in the case of illusions and hallucinations).²¹

Thus, in order to be emotionally rational, or at least to avoid recalcitrant emotions when necessary, we again must retreat to more indirect means like changing our setting, meditating, or going for a walk. In principle, of course, reasoning could also cause our emotions (as well as our sensations, or perceptual experiences) to change directly, i.e., without a mediating judgment. But that would not be rational determination, but much more like a causation by accident: ‘[i]n principle, anything can cause anything’ (Owens 2000, 102).

12.6 SUMMARY

In this chapter, I have reconciled two claims which seemed to be unreconcilable to philosophers who argue that we require a form of direct control over our attitudes in order to account for the special kind of agency we exercise over our own attitudes. I spelled out the puzzle throughout 12.1 and 12.2. The first claim is that the only form of control we have over our attitudes is indirect. The second claim is that we have a special kind of control over our attitudes which others lack over our attitudes for conceptual reasons. They seemed hard to reconcile because,

²¹ One objection considering which would bring me too far afield from the topic of this chapter claims that emotions are a species of beliefs that are especially susceptible to recalcitrance. I cannot rule out this possibility here, and there is much more to say about these issues, of course. Thanks to Konstantin Weber for pointing out this possible reply to me.

in principle, it seems that we can imagine any form of indirect control as being exercised over our own mind, and the mind of others. After all, if I influence my mind by performing a certain action, why cannot anyone else, in principle, also influence my mind by performing the same type of action? By solving this philosophical puzzle, I took another motivation for spelling out forms of direct mental control besides the idea that we need to assume such a control in order to explain mental responsibility.

It turned out that this last intuition is wrong. There are types of action by means of which we can only influence our own mind. We see this as soon as we note that some actions can result in realizing their constitutive aim, and that there is a type of reasoning that constitutively aims at coming to a conclusion, or determining one's attitudes in a rational way (12.3). I have argued, by means of thought experiments (12.4), that no one else could possibly determine our own attitudes in a rational way by reasoning to them. This is something agents can do only for themselves. I pointed out (in 12.2) that the reason for this is not the trivial reason that everyone can only exercise their own agency. Rather, the reason is that the relation of non-deviant causation that I wish to capture by the idea of causation in a rational way can only occur between the reasoning of a person and the attitudes of this person.

I have furthermore rejected the objection that I did not identify a special form of control because rational determination is nothing we control directly: it is sufficient that we have indirect control over rationally determining our attitudes in order to have a special kind of control over our own attitudes nobody else has (12.5.1). The control we exercise is *to rationally determine our attitudes by reasoning to them*. I finally considered the connection between rationality and reasoning that was brought into view by Broome. It turned out that while reasoning is often a means to satisfy requirements of (especially theoretical) rationality, there are often more effective means to do so (12.5.2). These more effective means, or more indirect strategies, seem to be not only especially called for when determining our intentions, but also when we need to take charge of our emotional life (12.5.3).

Conclusion

The aim of this inquiry was to critically examine *The easy solution* and to thereby provide a plausible account of mental responsibility. We saw that the main worry with *The easy solution* is that our attitudes can be held for reasons – i.e., are rationally evaluable. This is why, in chapter 2, I spelled out the problem of mental responsibility as a problem about understanding our responsibility to the standards of rationality. It seemed that there is no form of control that could explain our responsibility to these standards. And yet it seems that we are often criticized or blamed for failing to comply with them. In order to resolve this problem, I proposed, in chapter 3, to look closer at the status of rationality, i.e., at the status of reasons for attitudes.

Throughout part II, I assumed a constitutivist account of reasons for attitudes. That is, I understood reasons for attitudes as object-given. I have argued that this assumption leads to a challenge of understanding the normative force of reasons for attitudes. I have spelled out this challenge by discussing two arguments for mental nihilism, i.e., for the claim that there are no reasons for attitudes at all. If reasons for attitudes are not practical (like reasons for action), but rather object-given, then it is, *prima facie*, hard to see how we blame people for non-compliance with these reasons (chapters 4 and 5). Furthermore, cases in which we ought to cause non-compliance with object-given reasons throw further doubt on their status as reasons (chapter 6). If one of these two arguments was sound, and all reasons for attitudes were object-given, then there would be no reasons for attitudes. Consequently, it would be false that we always ought to respond correctly to object-given considerations – i.e., it would be false that we ought to be rational. This is the threat from mental nihilism.

The proponent of *The easy solution* could try to endorse nihilism: if ‘reasons’ for attitudes turn out not to deserve their status as reasons, then we thereby rid ourselves of the main problem for this solution. Yet I have argued that we should not endorse such a radical claim without hesitation. The arguments that called into doubt the very existence of reasons for attitudes had their weak spots. First, we sometimes seem to respond to others for their rational failures in ways we would only respond to responsible beings – with reactions like distrust or other attitudes that mark the impairment of an interpersonal relationship. This seems to allow for a kind of blameworthiness for rational failures, although the blame is quite distinct from the passionate blame that is sometimes appropriate when someone violates a requirement of morality. Secondly, we can make sense of cases in which we ought to cause ourselves not to comply with the standards of rationality and yet end up blameworthy for this non-compliance – e.g.,

we are worthy of distrust due to the resulting irrationality. We could make this intelligible by comparing the criticizability that is implied by someone's irrationality with the criticizability that is implied by someone's vices. My reactions to the threat from mental nihilism allowed us to distinguish two different senses of 'ought' and 'reason' that helped us make sense of the intuitive plausibility of the two arguments that gave rise to the threat by formulating sound versions of the argument that do not conclude in mental nihilism, but rather in acknowledging that standards of rationality must be understood differently from requirements of action (cf. 'Summary and Conclusion of Part II').

At the end of part II, I thus proposed that we should closer examine the relevant forms of criticizability or blameworthiness in order to better understand the status of reasons for attitudes. Doing so might, I suggested, help us see what, if any, form of control is implied by the kind of responsibility that is presupposed by this criticizability or blameworthiness. However, before I turned to this closer examination, I questioned one of my assumptions that lead us to the threat of mental nihilism: I called into question mental constitutivism by considering the prospects of mental pragmatism throughout part III. I have argued that, contrary to what some contemporary pragmatists claim, pragmatism implies voluntarism (chapter 7). I then considered different versions of voluntarism and rejected them all (chapters 8 and 9). I argued, first, that we cannot *form* attitudes at will because any formation of an attitude is either not direct or not active. Secondly, I have argued that we cannot believe, desire, feel, or intend at will. The first reading of this position assumes that attitudes are themselves active. I argued that this assumption is wrong as long as we understand 'active' in the ordinary way as involving actions. The second reading of this position does not make such an assumption. However, I have argued that without this assumption, it is hard to see how attitudes could ever be controlled in the same way as actions can be controlled.

Since pragmatism could not provide us with a satisfying account of reasons for attitudes, we were left again with the challenge from part II: In order to understand how we are responsible for (non-)compliance with object-given reasons, we should first understand the status of these reasons. We saw in part II that we could understand this status if we focus on how criticism and blame is appropriate for non-compliance with these reasons. Thus, the purpose of part IV was to understand this kind of blame, and how it differs from the blame that is appropriate only if our attitudes were under our indirect control. By spelling out both forms of blame, and the kinds of responsibility presupposed by each, we reached a satisfying account of mental responsibility that allowed us to say that *The easy solution* is plausible only if we assume a

narrow reading of ‘responsible.’ Yet under this reading, *The easy solution* is true, though under a broader understanding of ‘responsible,’ it is false.

To reach this *hybrid account of mental responsibility*, I started out distinguishing direct answerability from historical responsibility from attitudes (chapter 10). I have argued that both forms of responsibility are distinct and cannot be reduced to one another insofar as both are grounded in different kinds of reasons: the kinds of blameworthiness that correspond to these forms of responsibility presuppose different kinds of norm-violations (violating standards of rationality versus violating requirements of action). The distinctness of both types of responsibility gave rise to the worry that answerability is, after all, no form of responsibility. I have argued that we can meet this worry if we acknowledge the ambiguity in ‘ought’ and ‘reason’ that I pointed out at the end of part II. Instead of being worried by this ambiguity, we should acknowledge that there are two such distinct concepts and corresponding forms of blameworthiness and responsibility. This helps us to make sense of our intuitions in cases of conflict where subjects violate one type of norm but conform to the other. I finally (chapter 11) gave more substance to the distinction between the two types of responsibility by arguing that answerability is indeed a form of responsibility: otherwise we would not be able to explain our blameworthiness for involuntary disrespect. I distinguished the relevant senses of responsibility by associating them with the relevant blaming-responses: certain passionate responses are appropriate only if the person violated a requirement of mental self-management, but other responses, like distrust or judging that one’s relationship to the person is impaired, and behaving accordingly, do not presuppose indirect control.

My account of mental responsibility is, I think, faced with two main worries. The first is that, since I grant that we sometimes wrong others involuntarily in such a way that we owe them an apology, passionate responses like resentment and indignation should be appropriate responses to such wrongs even in cases where we lacked indirect control over our attitudes. This worry can be seen as giving more space to the involuntarist face of responsibility than I allow it to have: even passionate blame can be legitimate without control.

The second worry comes from mental compatibilism, and thereby wants to restrict the involuntarist aspect of my solution. It claims that I did not establish mental involuntarism, i.e., the claim that understanding mental responsibility does require us to assume that we control our attitudes directly *in any sense* of ‘direct control.’ I have mentioned these worries and reacted to them before, but I deem it to be helpful to summarize my replies in this conclusion. After I have done so, I will finally point out how further research on mental normativity is naturally motivated by my inquiry.

My reply to the first worry is twofold. First, I have argued throughout chapter 11 that this worry fades if we consider the cases in which we lack indirect control over attitudes by which we wrong others. In such cases, it seems rational for the passion of our emotions to fade. Secondly, however, I have provided a backup account for the hybrid account of mental responsibility that does not rely on this concrete proposal (cf. ‘Introduction to Part IV’). The backup account claims that *a different set of reactive attitudes* is appropriate to those who violate a standard of rationality than to those who violate a practical requirement of attitudinal management. The distinction between the two sets of attitudes needs not be drawn by the passionate element involved in the blaming-responses. The claim of this back-up account is weaker, and thus hopefully plausible to more people, than my concrete proposal about the nature of the sets of reactive attitudes. If there are two such different sets of reactive attitudes (which of course would not be mutually exclusive, but overlap with one another), then we can thereby define different types of blameworthiness and corresponding forms of responsibility. This backup account suffices to establish a version of the hybrid account of mental responsibility, and thus a reading under which *The easy solution* is true.

The second worry is that I did not establish mental involuntarism. Rather, my arguments might also be endorsed by a mental compatibilist (see chapter 2.6 on mental compatibilism). After all, I did not provide an extensive survey of accounts of direct non-voluntary control. My reason for this omission was that I deemed a different approach more promising: the approach of understanding mental responsibility by understanding the status of the standards of rationality. This approach seemed and proved to be promising because *The easy solution* was too simple an explanation of mental responsibility because attitudes can be, in contrast to mere sensations, held for reasons. This fact implies that we are, in a sense, directly responsible for them. Otherwise the status of these reasons as reasons would be mysterious.

I have argued that what grounds our responsibility to rational standards is that we are *answerable* for our attitudes, i.e., that they are responsive to reasons in such a way that we can be appropriately asked for our reasons for holding these attitudes. Understanding how answerability is a form of direct responsibility for attitudes is, I claim, sufficient for understanding how we are responsible to the standards of rationality. That is, it is sufficient for understanding how various blaming-responses are appropriate reactions to a person’s attitude even though this attitude was not under the person’s indirect control. If we *stipulate* a concept of ‘direct non-voluntary control’ that is co-extensive with an attitude’s reasons-responsiveness, then I have no issues with granting that our attitudes are, in this sense, under our ‘direct control.’ With such a

stipulative use of ‘direct non-voluntary control,’ the distinction between mental involuntarism and mental compatibilism would just collapse.

However, I did not deny that there might be other philosophical motivations for spelling out a concept of direct non-voluntary control that *adds* something to the idea that our attitudes are responsive to reasons. For example, we might think that in order to understand how agency is possible in world dominated by natural laws, we have to assume that not only our actions, but also the attitudes that have causal influence on them are somehow ‘under our direct control’ (cf. Hieronymi ms). Nothing I have said contradicts an argument that we need such a notion of control in order to solve other philosophical problems. My claim is merely that we need not assume such a notion of control for solving the problem of mental responsibility (which is the denial of what mental compatibilism, as defined in chapter 2.6, claims). In chapter 12, I have furthermore taken some motivation of spelling out such a notion of control in order to make sense of the idea that we are actively related to our own attitudes in a way we are not related to the attitudes of other people. For we can make sense of this idea merely by acknowledging that we have a specific form of indirect control over our own mind. We can thus preserve an intuitive self-other asymmetry even if we deny direct mental control over our own attitudes.

One important claim that was central to my investigation is that in order to legitimately call a consideration a *reason* for a response, it must be possible to be blameworthy for non-compliance with the kinds of reasons to which this consideration is supposed to belong.¹ If we thus want to preserve the idea that reasons for attitudes deserve their status as reasons, we need to think about blameworthiness and criticizability, i.e., about responsibility. I thus propose here that we can understand reasons in terms of responsibility. This is at odds with the ‘reasons first’-program in contemporary epistemology and metaethics, which claims that we can understand all normative phenomena in terms of reasons.² I am not strictly opposed to this program. However, I deem it to be dubious why we should only pursue an understanding of a concept in one direction. If responsibility can be understood in terms of reasons, then pursuing an understanding of reasons in terms of responsibility might gain us with interesting insights as well. More precisely, understanding the normative force of reasons in terms of different forms of blameworthiness or criticizability might help us to understand and differentiate between different types of reasons. If I am right that reasons deserve their status only if they can give rise to

¹ Cf. the first premise of the arguments from blameworthiness in chs. 4 and 5.

² Classical proponents of the program are, according to Sosa and Sylvan (2018), Scanlon (1998), Parfit (2011), Schroeder (2007), and Skorupski (2011). Critical discussions are to be found in Broome 2004; Crisp 2006; Thomson 2008; Väyrynen 2011; Wedgwood 2015; 2017.

blameworthiness or criticizability, then understanding reasons requires us to get clear about these types of blameworthiness and criticizability.

I provided some first outlines of this project within the present book. Yet I only briefly considered the phenomenology of blame for different types of rational failures. How do we hold each other responsible for epistemic failures and corresponding epistemic vices, like dogmatism or gullibility? And how do we sometimes blame someone for being akratic, or credit someone for being enkratic, i.e., intending what they think they ought to do? What kinds of responses are appropriate for various emotional vices insofar as they exhibit irrationality, like intemperance or impatience? If my assumption that there is a conceptual connection between reasons and blameworthiness is correct, then these questions are not only intrinsically of interest and helpful for understanding the nature and normativity of interpersonal relationships. Rather, they are essential for any account of reasons. Understanding reasons requires us to understand responsibility. However, a full understanding of reasons in terms of responsibility is beyond the scope of this inquiry.

Literature

- Adams, Robert M. 1985. 'Involuntary Sins', *The Philosophical Review* 94, 3-31.
- Adler, Jonathan. 2002. *Belief's Own Ethics*, Cambridge, MA: MIT Press.
- Alston, William P. 1988. 'The Deontological Conception of Epistemic Justification', *Philosophical Perspectives* 2, 257-299.
- Alvarez, Maria. 2010. *Kinds of Reasons. An Essay in the Philosophy of Action*, NY: OUP.
- 2017. 'Reasons for Action: Justification, Motivation, Explanation', *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), ed. by Edward N. Zalta, URL = <<https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/>>.
- 2018. 'Reasons for action, acting for reasons, and rationality', *Synthese* 195, 3293-3310.
- Anscombe, G. E. M. 1957. *Intention*, Cambridge, MA: HUP.
- Audi, Robert. 2001. 'Doxastic voluntarism and the ethics of belief', in: Matthias Steup (dd.), *Knowledge, truth, and duty: Essays on epistemic justification, responsibility, and virtue*, NY: OUP, 63-76.
- Basu, Rima. 2019. 'The wrongs of racist beliefs', *Philosophical Studies* 176, 2497-2515.
- Basu, Rima / Schroeder, Mark 2019. 'Doxastic Wronging', in: *Pragmatic Encroachment in Epistemology*, ed. by Kim, Brian / McGrath, Matthew, London: Routledge, 181-205.
- Beckermann, Ansgar. 2013. *Glaube*, Berlin: de Gruyter.
- Bennett, Jonathan. 1990. 'Why Is Belief Involuntary?' *Analysis* 50, 87-107.
- Boyle, Matthew 2011. "'Making up Your Mind" and the Activity of Reason', *Philosopher's Imprint* 11, 1-24.
- Broome, John. 2004. 'Reasons', in: Wallace, R. Jay / Smith, Michael / Scheffler, Samuel / Pettit, Philipp (eds.), *Reason and Value: Essays on the Moral Philosophy of Joseph Raz*, NY: OUP, 28-55.
- 2007. 'Does Rationality Consist in Responding Correctly to Reasons?' *Journal of Moral Philosophy* 4, 349-74.
- 2013. *Rationality Through Reasoning*, NY: OUP.
- Calhoun, Cheshire. 1989. 'Responsibility and Reproach', *Ethics* 99, 389-406.
- Chan, Timothy (ed.). 2013. *The Aim of Belief*, NY: OUP.
- Chrisman, Matthew. 2008. 'Ought to believe', *The Journal of Philosophy* 105, 346-370.
- 2016. 'Epistemic Normativity and Cognitive Agency', *Noûs*, DOI: 10.1111/nous.12184.
- fc. 'Believing as We Ought and the Democratic Route to Knowledge', in: Schmidt/Ernst fc.
- Clifford, William K. 1877. 'The Ethics of Belief', in: *The Ethics of Belief. And other Essays*, NY: Prometheus Books 1999, 70-96.
- Coates, D. Justin. fc. 'The Ethics of Blame: A Primer', in: Schmidt/Ernst fc.
- Coates, D. Justin / Tognazzini, Neal. 2013. *Blame. Its Nature and Norms*, NY: OUP.
- Cohen, Stewart. 1984. 'Justification and Truth', *Philosophical Studies* 46, 279-95.
- Crawford, Lindsay. fc. 'Relationships and Reasons for Belief', in: Schmidt/Ernst fc.
- Crisp, Roger. 2006. *Reasons and the Good*, NY: OUP.
- Davidson, Donald. 1963. 'Actions, Reasons, and Causes', *The Journal of Philosophy* 60, 685-700.
- Dawkins, Richard. 2006. *The God Delusion*, London: Bantam Press.
- Descartes, René. 1649. *The Passions of the Soul*, in: *The Philosophical Writings of Descartes: Volume I*, transl. by Cottingham et al., Cambridge: CUP, 1985.
- Döring, Sabine. 2010. 'Why be emotional?', in: Goldie, Peter (ed.): *The Oxford Handbook of Philosophy of Emotion*, NY: OUP, 283-301.
- fc. 'How safe should we feel? On the ethics of fear in the Public sphere', in: Schmidt/Ernst fc.

- D'Oro, Guiseppina / Sandis, Constantine (eds.). 2013. *Reasons and Causes: Causalism and Non-causalism in the Philosophy of Action*, London: Palgrave-Macmillan.
- Epictetus. *Encheiridion*, Mineola: Dover Publications.
- Ernst, Gerhard. 2008. *Die Objektivität der Moral*, Paderborn: mentis.
- Feldman, Richard. 2000. 'The Ethics of Belief', *Philosophy and Phenomenological Research* 60, 667-695.
- 2007. 'Reasonable Religious Disagreements,' in *Philosophers without God: Meditations on Atheism and the Secular Life*, ed. by Louise Antony, NY: OUP, 194-214.
- Fingarette, Herbert. 1969. *Self-Deception*, London: Routledge.
- Fischer, John / Ravizza, Mark. 1998. *Responsibility and Control. A Theory of Moral Responsibility*, Cambridge: CUP.
- Fischer, John / Tognazzini, Neal. 2009. 'The Truth about Tracing', *Noûs* 43, 531-556.
- Foley, Richard. 1993. *The Theory of Epistemic Rationality*, Cambridge, MA: HUP.
- Foot, Philippa. 1958: 'Moral Arguments', in: *Mind* 67, 502-513.
- 2001. *Natural Goodness*, NY: OUP.
- Frankfurt, Harry G. 1969. 'Alternate Possibilities and Moral Responsibility', *The Journal of Philosophy* 66, 829-839.
- Fumerton, Richard. 1995. *Metaepistemology and Skepticism*, Lanham: Rowman & Littlefield.
- Ginet, Carl. 2001. 'Deciding to Believe', in: Steup 2001, 63-75.
- Glüer, Kathrin / Wikforss, Åsa. 2018. 'Reasons for Belief and Normativity', in: Star, Daniel (ed.), *The Oxford Handbook of Reasons and Normativity*, NY: OUP, <https://doi.org/10.1093/oxfordhb/9780199657889.013.26>.
- Graham, Peter A. 2014. 'A Sketch of a Theory of Moral Blameworthiness', *Philosophy and Phenomenological Research* 88, 388-409.
- Hacker, Peter M. S. 1990. *Wittgenstein – Meaning and Mind*, Oxford: Blackwell.
- Hare, Richard M. 1952: *The Language of Morals*. NY: OUP (reprinted 1964).
- 1981. *Moral Thinking. Its Levels, Method and Point*. Oxford: Clarendon Press.
- Harman, Gilbert. 1986. *Change in View: Principles of Reasoning*. Cambridge, MA: MIT Press.
- Hazlett, Allan. 2013. *A Luxury of the Understanding. On the Value of True Belief*, Oxford: OUP.
- Heuer, Ulrike. 2018. 'The Relevance of the Wrong Kind of Reasons', in: McHugh, C. / Way, J. (eds.): *Normativity: Epistemic and Practical*, NY: OUP.
- Hieronymi, Pamela. 2004. 'The Force and Fairness of Blame', *Philosophical Perspectives* 18, 115-148.
- 2006. 'Controlling Attitudes', *Pacific Philosophical Quarterly* 87, 45-74.
- 2008. 'Responsibility for believing', *Synthese* 161, 357-373.
- 2009. 'Two Kinds of Agency', in: *Mental Actions*, ed. by O'Brian, Lucy / Soteriou, Matthew, NY: OUP, 138-168.
- 2014. 'Reflection and Responsibility', *Philosophy and Public Affairs* 42, 3-41.
- 2019. 'I'll Bet You Think This Blame is About You', in: *Oxford Studies in Agency and Responsibility V: Themes from the Philosophy of Gary Watson*, ed. by Coates, D. Justin / Tognazzini, Neal, NY: OUP, 60-87.
- ms. *Minds that Matter*, unpublished manuscript (spring 2019 version).
- Huemer, Michael 2001. *Skepticism and the Veil of Perception*, Lanham: Rowman & Littlefield.
- Hussain, Nadeem J. Z. 'The Requirements of Rationality'. Unpublished manuscript, Stanford University, August 2007, version 2.4.
- Hyman, John. 2015. *Action, Knowledge, and Will*, NY: OUP.
- James, William. 1886. 'What is An Emotion?', *Mind* 9, 188-205.
- 1896. 'The Will to Believe', in *The Will to Believe. And other Essays in popular Philosophy, and Human Immortality*, Dover: Dover publishing 1956, 1-31.
- Kavka, Gregory S. 1983. 'The Toxin Puzzle', *Analysis* 43, 33-36.

- Kearns, Stephen / Star, Daniel. 2009. 'Reasons as Evidence', in Shafer Landau (ed.), *Oxford Studies in Metaethics*, vol. 4, NY: OUP, 215-42.
- Kelly, Thomas. 2002. 'The rationality of belief and some other propositional attitudes', *Philosophical Studies* 110, 163-196.
- 2003. 'Epistemic Rationality as Instrumental Rationality: A Critique', *Philosophy and Phenomenological Research* 66, 612-640.
- Kiesel, Dagmar / Schmidt, Sebastian: 'Freier Wille, Personale Identität und epistemische Ungewissheit', in: Kiesel, D. / Ferrari, C.: *Willensfreiheit*, Frankfurt am Main: Klostermann, 221-258.
- Kiesewetter, Benjamin. 2015. 'Instrumental Normativity: In Defense of the Transmission Principle', *Ethics* 125, 921-46.
- 2016. 'You Ought to ϕ Only if You May Believe that You Ought to ϕ ', *The Philosophical Quarterly* 66, 760-782.
- 2017. *The Normativity of Rationality*, Oxford: OUP.
- ms. 'Are Epistemic Reasons Normative?', unpublished manuscript, July 2019 version.
- Kolodny, Niko. 2005. 'Why Be Rational?' *Mind* 114, 509-63.
- Korcz, Keith Allen. 2019. 'The Epistemic Basing Relation', *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), ed. by Edward N. Zalta, URL = <<https://plato.stanford.edu/archives/fall2019/entries/basing-epistemic/>>.
- Kornblith, Hilary. 1983. 'Justified Belief and Epistemically Responsible Action', *Philosophical Review* 92, 33-48.
- Korsgaard, Christine. 1992. 'Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations', *Philosophical Perspectives* 6, 305-332.
- Kruse, Andrea. 2015. 'Why doxastic responsibility is not based on direct doxastic control', *Synthese*, 1-32.
- Levy, N. / Mandelbaum, E. 2014. 'The Powers that bind: doxastic voluntarism and epistemic obligation,' in: Jonathan Matheson (ed.), *The Ethics of Belief*. New York: OUP, 12-33.
- Lindner, Martina. fc. 'Implications of the Debate on Doxastic Voluntarism for Clifford's Ethics of Belief', in: Schmidt/Ernst fc.
- Littlejohn, Clayton / Turri, John (eds.). 2014. *Epistemic Norms. New Essays on Action, Belief, and Assertion*, NY: OUP.
- Lord, Errol. 2018. *The Importance of Being Rational*, NY: OUP.
- fc. 'Suspension of Judgment, Rationality's Competition, and the Reach of the Epistemic', in: Schmidt/Ernst (eds.).
- Maguire, Barry. 2018. 'There Are No Normative Reasons for Affective Attitudes', *Mind* 127, 779-805.
- McCormick, Miriam S. 2011. 'Taking control of belief', *Philosophical Explorations* 16, 169-183.
- 2015. *Believing Against the Evidence. Agency and the Ethics of Belief*, London: Routledge.
- 2017. 'Rational Hope', *Philosophical Explorations* 20, 127-141.
- 2018. 'Responding to Skepticism about Doxastic Agency', *Erkenntnis* 83, 627-645.
- fc. 'The Value of a Free and Wandering Mind', in: Schmidt/Ernst fc.
- McCormick, Miriam / Schleifer, Michael. 2006a. 'Responsibility for Beliefs and Emotions', *Paideusis: Journal of the Canadian Philosophy of Education Society* 15, 75-85.
- 2006b. 'Are We Responsible for Our Emotions and Moods?', *Thinking: The Journal of Philosophy of Children* 18, 15-21.
- McDowell, John. 1994. *Mind and World*, Cambridge, MA: HUP.
- McHugh, Conor. 2013a. 'Epistemic Responsibility and Doxastic Agency', *Philosophical Issues* 23, 132-156.
- 2013b. 'Normativism and Doxastic Deliberation', *Analytic Philosophy* 54, 447-465.
- 2014a. 'Attitudinal Control', *Synthese*, 1-18.

- 2014b. ‘Exercising Doxastic Freedom’, *Philosophy and Phenomenological Research* 88, 1-37.
- McKenna, Michael. 2016. ‘Reasons-Responsive Theories of Freedom’, in: *The Routledge Companion to Free Will*, ed. by K. Timpe et al., London: Routledge, 27-40.
- Meiland, Jack W. 1980. ‘What Ought we to Believe? Or the Ethics of Belief Revisited’, *American Philosophical Quarterly* 17, 15-24.
- Mele, Alfred. 1990. ‘Errant Self-Control and the Self-Controlled Person’, *Pacific Philosophical Quarterly* 71: 47-59.
- Meylan, Anne. 2013. *Foundations of an Ethics of Belief*, Frankfurt: ontos.
- 2017. ‘The Consequential Conception of Doxastic Responsibility’, *Theoria* 83, 4-28.
- 2018. ‘The Reasons-Responsiveness Account of Doxastic Responsibility and the Basing Relation’, *Erkenntnis* 84, 877-893.
- Montmarquet, James. 1993. *Epistemic Virtue and Doxastic Responsibility*, Lanham: Rowman and Littlefield.
- Moran, Richard. 1988. ‘Making up your mind: Self-Interpretation and Self-Constitution’, *Ratio* 1, 135-151.
- 2001. *Authority and Estrangement. An Essay on Self-Knowledge*, Oxford: PUP.
- Mourelatos, Alexander. 1978. ‘Events, Processes, and States’, *Linguistics and Philosophy* 2, 415-34.
- Nagel, Thomas. 1986. *The View from Nowhere*, NY: OUP.
- Neu, Jerome. 2009. ‘An Ethics of Emotion?’, in: Peter Goldie (ed.), *The Oxford Handbook of Philosophy of Emotion*, NY: OUP.
- Nottelmann, Nikolaj. 2006. ‘The Analogy Argument for Doxastic Voluntarism’, *Philosophical Studies* 131, 559-582.
- Levy, Neil. 2007. ‘Doxastic Responsibility’, *Synthese* 155, 127 – 155.
- Oakley, Justin. 1992. *Morality and the Emotions*, London: Routledge.
- Olfert, Christiana. fc. *The Important Problem. Pyrrhonian Skepticism as a Way of Life*, NY: OUP.
- Owens, David J. 2000. *Reason without Freedom. The problem of epistemic normativity*, London: Routledge.
- 2003. ‘Does Belief Have an Aim?’, reprinted in his *Normativity and Control*, 51-68.
- 2017a. *Normativity and Control*, Oxford: OUP.
- 2017b. ‘Value and Epistemic Normativity’, in his *Normativity and Control*, 87-109.
- Parfit, Derek. 2001. ‘Rationality and Reasons’, in: *Exploring Practical Philosophy*, ed. by Egonsson et al., Burlington, VT: Ashgate, 17-39.
- 2011. *On What Matters. Volume I*, NY: OUP.
- Papineau, David. 1999. ‘Normativity and Judgment’, in his *The Roots of Reason*, NY: OUP 2006, 6-38.
- 2013. ‘There Are No Norms of Belief’, in: Chan 2013, 64-79.
- Pascal, Blaise. 1670. *Pensées*, Seattle: Pacific Publishing Studios 2011.
- Peels, Rik. 2017. *Responsible Belief*, NY: OUP.
- Pettit, Philip. 2004. ‘Hope and Its Place in Mind’, *Annals of the American Academy of Political and Social Science* 592, 152-165.
- Pink, Thomas. 2009. ‘Reason, voluntariness and moral responsibility’, in *Mental Actions*, ed. by O’Brian, L. / Soteriou, M., Oxford: OUP, 95-120.
- Price, Carolyn. 2015. *Emotion*, Cambridge: Polity Press.
- Price, Henry H. 1954. ‘Belief and Will’, *Proceedings of the Aristotelian Society, Supplementary Volumes* 28, 1-26.
- Rabinowicz, Wlodek / Rønnow-Rasmussen, Toni. 2004. ‘The Strike of the Demon: On Fitting Pro-Attitudes and Value’, *Ethics* 114, 391-423.

- Raz, Joseph. 2005a. 'Instrumental Rationality: A Reprise', *Journal of Ethics and Social Philosophy* 1, 1-19.
- 2005b. 'The Myth of Instrumental Rationality', *Journal of Ethics and Social Philosophy* 1, 2-28.
- 2011. *From Normativity to Responsibility*, NY: OUP.
- Reisner, Andrew. 2008. 'Weighing pragmatic and evidential reasons for belief', *Philosophical Studies* 138, 17-27.
- 2009. 'The possibility of pragmatic reasons for belief and the wrong kind of reasons problem', *Philosophical Studies* 145, 257-272.
- Reisner, Andrew / Steglich-Petersen, Asbjørn (eds.). 2011. *Reasons for Belief*, Cambridge: CUP.
- Rinard, Susanna. 2015. 'Against the New Evidentialists', *Philosophical Issues* 25, 208-223.
- 2017. 'No Exception for Belief', *Philosophy and Phenomenological Research* 94, 121-143.
- Roberts, Tom. 2015. 'Emotional Regulation and Responsibility', *Ethical Theory and Moral Practice* 18, 487-500.
- Rosen, Gideon. 2004. 'Skepticism about Moral Responsibility', *Philosophical Perspectives* 18, 295-313.
- Scanlon, Thomas M. 1998. *What We Owe to Each Other*, Cambridge, MA: HUP.
- 2008. *Moral Dimensions. Permissibility, Meaning, Blame*, Cambridge, MA: Belknap Press.
- Schmidt, Sebastian. 2016. 'Können wir uns entscheiden, etwas zu glauben? Zur Möglichkeit und Unmöglichkeit eines doxastischen Willens', *Grazer Philosophische Studien* 93, 571-582.
- 2017. 'Why We Should Promote Irrationality', *Grazer Philosophische Studien* 94, 605-615.
- fca. 'Introduction', in: Schmidt/Ernst fc.
- fcb. 'Mental Responsibility, Object-given Reasons, and Blame', in: Schmidt/Ernst fc.
- Schmidt, Sebastian / Ernst, Gerhard. fc. *The Ethics of Belief and Beyond. Understanding Mental Normativity*, London: Routledge.
- Schroeder, Mark. 2007. *Slaves of the Passions*, NY: OUP.
- Skorupski, John. 2011. *The Domain of Reasons*, NY: OUP.
- Scott-Kakures, Dion. 1994. 'On Belief and the Captivity of the Will', *Philosophy and Phenomenological Research* 54, 77-103.
- Setiya, Kieran. 2007. *Reasons without Rationalism*, Princeton: PUP.
- 2013. 'Epistemic Agency: Some doubts', *Philosophical Issues* 23, 179-198.
- Sextus Empiricus. *Outlines of Pyrrhonism*, Cambridge, MA: HUP.
- Shoemaker, Sidney (ed.). 1963. *Self-Knowledge and Self-Identity*, Ithaca: Cornell University Press.
- Shah, Nishi. 2006. 'A New Argument for Evidentialism', *The Philosophical Quarterly* 56, 481-498.
- 2008. 'How Action Governs Intention', *Philosophers Imprint* 8, 1-19.
- Sher, George. 2006. *In Praise of Blame*, NY: OUP.
- 2009. *Who Knew? Responsibility Without Awareness*, NY: OUP.
- Skipper, Matthias / Steglich-Petersen, Asbjørn. fc. 'Instrumental reasons for belief: elusive talk and elliptic properties', in: Schmidt/Ernst fc.
- Smith, Angela M. 2005. 'Responsibility for Attitudes: Activity and Passivity in Mental Life', *Ethics* 115, 236-271.
- 2013. 'Moral Blame as Moral Protest', in: *Blame*, ed. by Coates/Tognazzini, NY: OUP, 27-48.
- 2015a. 'Attitudes, Tracing and Control', *Journal of Applied Philosophy* 32, 115-132.
- 2015b. 'Responsibility as Answerability', *Inquiry* 58, 99-126.
- Solomon, Robert C. 2007. *True to Our Feelings*, NY: OUP.

- Sosa, Ernest / Sylvan, Kurt. 2018. 'The Place of Reasons in Epistemology', in: Star, Daniel (ed.), *The Oxford Handbook of Reasons and Normativity*, NY: OUP, 555-574.
- Steglich-Petersen, Asbjørn. 2009: 'Weighing the aim of belief', *Philosophical Studies* 145, 395-405.
- 2011. 'How to be a teleologist about epistemic reasons', in: Reisner/Steglich-Petersen 2011, 13-33.
- Steup, Matthias (ed.). 2001. *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, NY: OUP.
- Steup, Matthias. 2008. 'Doxastic Freedom', *Synthese* 161, 375-392.
- Stocker, Michael 1982. 'Responsibility Especially for Beliefs', *Mind* 91, 398-417.
- Stroud, Sarah. 2006. 'Epistemic Partiality in Friendship', *Ethics* 116, 498-524.
- Strawson, Peter F. 1962: 'Freedom and Resentment', *Proceedings of the British Academy* 48, 1-25.
- Sylvan, Kurt. ms. 'On Divorcing the Rational and the Justified', unpublished manuscript, August 2019 version.
- Tappolet, Christine. 2000. *Émotions et valeurs*, Paris: Presses Universitaires de France.
- 2016. *Emotions, Values, and Agency*, NY: OUP.
- Thomson, Judith J. 2008. *Normativity*, Chicago: Open Court.
- Tognazzini, Neal. fc. 'Silence and Salience. On Being Judgmental', in: Schmidt/Ernst fc.
- Väyrynen, Pekka. 2011. 'A Wrong Turn to Reasons?' in: Brady, Michael (ed.), *New Waves in Metaethics*, NY: Palgrave Macmillan.
- Vendler, Zeno. 1957. 'Verbs and Time'. *Philosophical Review* 66, 143-60.
- Wallace, R. Jay. 1994: *Responsibility and the Moral Sentiments*, Cambridge, Mass.: HUP.
- 2011. 'Dispassionate Opprobrium: On Blame and the Reactive Sentiments', in: *Reason and Recognition: Essays on the Philosophy of T. M. Scanlon*, NY: OUP.
- Watson, Gary. 1996. 'Two Faces of Responsibility', in: ders. (hg.), *Agency and Answerability*, NY: OUP 2004, 260-288.
- 2002. 'Volitional Necessities', in: ders. (hg.), *Agency and Answerability*, NY: OUP 2004, 88-122.
- Wagner, Verena. 2015. 'On the analogy of free will and free belief', *Synthese* 194, 2785–2810.
- Wedgwood, Ralph. 2002. 'The aim of belief', *Philosophical Perspectives* 36, 267-297.
- 2015. 'The Pitfalls of "Reasons"', *Philosophical Issues* 25, 123-143.
- 2017. *The Value of Rationality*, NY: OUP.
- White, Roger. 2005. 'Epistemic Permissiveness,' *Philosophical Perspectives* 19, 445-459.
- Williams, Bernard. 1970. 'Deciding to Believe', in his *Problems of the Self*, Cambridge: CUP 1973, 136-151.
- 1995. 'Internal Reasons and the Obscurity of Blame', in his *Making Sense of Humanity*, Cambridge: CUP, 35-45.
- Winters, Barbara. 1979. 'Believing at Will', *The Journal of Philosophy* 76, 243-256.
- Wolf, Susan. 1994. *Freedom Within Reason*, NY: OUP.
- 2010. *Meaning in Life and Why It Matters*, Princeton: PUP.
- Wrenn, Chase B. 2007. 'Why There Are No Epistemic Duties', *Dialogue* 46, 115-36.